A MIXED INTEGER-LINEAR PROGRAMMING MODEL FOR SOLVING
THE HYDROELECTRIC UNIT MAINTENANCE SCHEDULING PROBLEM

by

YUEHAO TANG

B.Sc. Wuhan University, China, 1999
M.A.Sc. Wuhan University, China, 2002

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(CIVIL ENGINEERING)

THE UNIVERSITY OF BRITISH COLUMBIA

April 2007
© Yuehao Tang, 2007
ABSTRACT

The unit maintenance scheduling problem is traditionally a challenging topic for electric utilities. In general, the system operators schedule the generating units for maintenance periodically. Research on the maintenance scheduling problem started in 1960's in the last century and has continued until today. Although significant contributions were made as a results of this continuous investigations, but many of those approaches were made in specific areas as different systems have their own special characteristics, such as system reliability considerations. According to the literature cited to-date, there is not one a successful case reported for maintenance scheduling of a pure hydro or hydro-dominated generating system.

This thesis outlines an optimization modelling approach that is based on the linear programming techniques to solve the maintenance scheduling problem for hydro systems. This research focused on the potential use of mixed-integer/linear programming algorithms to solve the maintenance scheduling problem for a large scale hydroelectric system. The objective was to determine the best “timing” for each unit outage in the system. This thesis reformulated an existing hydro scheduling model and concentrated on the set of constraints that could effectively satisfy the maintenance scheduling requirements.

The case studies carried out revealed that the most appropriate mixed-integer maintenance scheduling model would be formulated in such a way that it develops a maintenance schedule that ensures that the plants are operated in a way to maximize system efficiency and that head convergence have an influence on hydroelectric unit outage scheduling algorithm. Integer programming methods are usually considered suitable for maintenance scheduling problems, but in view of the complexities inherent in hydro models and in mixed-integer programming algorithms, it was found that it is very difficult to generalize one standardized approach to effectively handle this complex and large scale problem. Therefore, this research has focused on two main points. One focused on decreasing the number of possible binary variables in the model. The second focused on finding an appropriate algorithm that could approximate and reasonably simplify the computational process and transforms non-linear constraints into linear ones. It was also found that without specifically or delicately designing the maintenance scheduling constraints, the problem will be hard to solve even by the latest commercially available algorithms.
# TABLE OF CONTENTS

**ABSTRACT** .......................................................................................................................... ii

**TABLE OF CONTENTS** .......................................................................................................... iii

**LIST OF TABLES** .................................................................................................................. vi

**LIST OF FIGURES** ................................................................................................................ v

**ACKNOWLEDGEMENTS** ......................................................................................................... ix

**CHAPTER 1 INTRODUCTION** ................................................................................................. 1

1.1. Background ......................................................................................................................... 1

1.2. The Decision Marketing Environment ............................................................................... 4

1.2.1. Introduction to the deregulated electricity marketplace ................................................. 4

1.2.2. The role and state of B.C electric industry .................................................................. 4

1.3. Statement of the Problem ................................................................................................. 6

1.4. Goal, Objective and Study Approach ............................................................................... 9

1.5. Organization of this thesis ............................................................................................... 9

**CHAPTER 2 LITERATURE REVIEW** ...................................................................................... 10

2.1. General Overview ............................................................................................................. 10

2.1.1. Broad Review of Maintenance Research Development ............................................. 10

2.1.2. Common Maintenance Models and Their Applicability .......................................... 12

2.1.3. Generalized Maintenance Frameworks ...................................................................... 14

2.2. Maintenance Scheduling in Electricity Utilities .............................................................. 15

2.2.1. General Introductions ................................................................................................. 15

2.2.2. Linear / Integer Based Programming ....................................................................... 17

2.2.3. Reliability Centred Analysis ...................................................................................... 19

2.2.4. Dynamic programming model ................................................................................... 26

2.2.5. Decomposition / Hierarchical Model ........................................................................ 27

2.2.6. Fuzzy Logic or Knowledge-based model .................................................................. 30

2.2.7. Heuristic / Genetic Based Algorithms ....................................................................... 32

2.2.8. Miscellaneous Coordination Techniques ................................................................... 36

2.2.9. Game Theory ............................................................................................................. 37

2.3. Maintenance Optimization Models in Other Industries ....... ................................. 38

**CHAPTER 3 MODELING METHODOLOGY AND SOLUTION ALGORITHM** .............. 41

3.1. Introduction ....................................................................................................................... 41
3.2. Mixed-Integer Programming Model

3.2.1. GOMMS Water Balance Constraint

3.2.2. Load-Resource Balance, Spinning Reserve and Market Limit Constraints

3.2.3. Plant Generation and Turbine Flow Constraints

3.2.4. Maintenance Constraints

3.2.5. Reliability Requirements

3.2.6. Rotational Energy Requirement

3.2.7. Synchronous Condenser

3.2.8. Objective Function

3.2.9. Other Constraints or Considerations

3.3. Simple Linear Relaxation Scheme

3.3.1. Brief Review of Linear Relaxation

3.3.2. Using Cutting-Plane Method to Solve the LP Relaxation Scheme

CHAPTER 4 CASE STUDIES AND RESULTS

4.1. Case Study Introduction

4.1.1. Reservoir Nature Inflows

4.1.2. Reservoir Forebay Elevations and Limits

4.1.3. System Resource Balance and Markets Activities

4.2. Numerical Studies and Comparisons

4.2.1. Single sub-time step results

4.2.1.1. Mixed integer programming model without head convergence scheme

4.2.1.2. Mixed integer programming model with head convergence scheme

4.2.1.3. Simple Linear Relaxation Scheme

4.2.2. Multiple sub-time step results

4.2.2.1. Mixed Integer Programming Model

4.2.2.2. Maintenance Shape Approach

4.2.2.3. Simple Linear Relaxation Scheme

4.3. Other Measures Concerning Solving Efficiency

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS

5.1. Summary

5.2. Future Development

BIBLIOGRAPHY

APPENDICES

Appendix A. Linear Interpolation Scheme

Appendix B. General Equations for Unit Maintenance Decision
LIST OF TABLES

Table 3-1: Combo example ........................................................................................................ 52
Table 3-2: parameter OutageTiming for unit 1 of MCA .................................................. 55
Table 3-3: parameter OutageTiming for unit 4 of MCA .................................................. 55
Table 3-4: Equality constraints by (3.32) ........................................................................ 56
Table 3-5: Operable Segments of MCA S\C Units .......................................................... 60
Table 4-1: Hydro plants and units’ information (MW) .................................................. 67
Table 4-2: illustrative Case 1 .......................................................................................... 87
Table 4-3: illustrative Case 2 .......................................................................................... 87
Table 4-4: Number of maintenance shapes .................................................................. 88
LIST OF FIGURES

Figure 1-1: BC Hydro Generation Map ................................................................. 1
Figure 1-2: Map of BC Hydro Service Areas ......................................................... 3
Figure 1-3: Sources of Electricity Supply in British Columbia (BC Hydro 2004) .......... 9
Figure 3-1: Illustrative Hydro Plant Generation Curve under certain combo and forebay ... 46
Figure 3-2: MCA HPG surface at combo 15 ......................................................... 47
Figure 3-3: GMS max outputs under different combo & forebay ......................... 49
Figure 3-4: The \( \theta \) grids of \( P \) with respect to \( FB \) and \( Q \) ................................. 50
Figure 3-5: the configuration of \( \lambda \) and its evaluation scheme .......................... 50
Figure 3-6: Relations of BC load, Net limits and Rotational energy ...................... 58
Figure 3-7: Changing tendency of slopes and intercepts ....................................... 58
Figure 3-8: S/C status and operable segments in HPG curve ............................... 59
Figure 4-1: GMS weekly inflows in 1964 .......................................................... 67
Figure 4-2: PCN weekly inflows in 1964 ............................................................ 68
Figure 4-3: MCA weekly inflows in 1964 ............................................................ 68
Figure 4-4: REV weekly inflows in 1964 ............................................................ 69
Figure 4-5: ARD weekly inflows in 1964 ............................................................ 69
Figure 4-6: GMS forebay trajectory and monthly limitations in 1964 ...................... 70
Figure 4-7: MCA forebay trajectory and monthly limitations in 1964 ...................... 71
Figure 4-8: ARD forebay trajectory and monthly limitations in 1964 ...................... 71
Figure 4-9: Load and resource balance plot (2008-2009 forecast load year) ........... 72
Figure 4-10: Weekly average price forecasts in Alberta and US markets ................. 73
Figure 4-11: The objective values and water head errors with respect to iteration .......... 76
Figure 4-12: Mixed-integer model without water head convergence ...................... 77
Figure 4-13: PCN forebay comparison in two models ......................................... 78
Figure 4-14: MCA forebay comparison in two models ......................................... 79
Figure 4-15: The objective values and water head errors with respect to iteration .......... 80
Figure 4-16: Mixed-integer model with water head convergence ......................... 81
Figure 4-17: GMS forebay by the simple linear relaxation scheme and MIP model .... 82
Figure 4-18: The objective values and water head errors with respect to iteration .......... 83
Figure 4-19: Simple linear relaxation model without water head convergence ........... 84
Figure 4-20: Mixed-integer model without water head convergence ........................ 86
Figure 4-21: the "Shape" of maintenance activity ................................................................. 88
Figure 4-22: Optimal values at different outage durations starting from March ..................... 90
Figure 4-23: The objective values and water head errors with respect to iteration .................. 91
Figure 4-24: Renewed maintenance scheduling in the consecutive manner ............................ 92
Figure 4-25: REV one unit outage cost .................................................................................. 93
Figure 4-26: Simple linear relaxation model without water head convergence ....................... 94
Figure 4-27: Division process of enumeration trees ................................................................. 96
Figure Appendix A-1: Linear interpolation of (3-16) to (3-17) .................................................... 108
Figure Appendix A-2: Approximation in generation surfaces (eqn. 3-13) ................................. 109
Figure Appendix B-1: Binary variables in the control the maintenance events ..................... 112
ACKNOWLEDGEMENTS

This thesis would not be finished without the guidance and support of Dr. Ziad Shawwash. His brilliant ideas and suggestions helped me understand and resolve all kinds of problems and give me a new way of thinking. His warm encouragements helped me explore new ideas and attempt new methods and techniques. His constant support and quiet attitude allowed completion of this research without too much pressure or tension. Even during his absence due to his illness, he still directed the research in time by email.

Special appreciation should also be made to Dr. Thomas K. Siu. He is the principal engineer at BC Hydro. His deep understanding of the mechanism and essence of BC Hydro system directly impacted how I have approached my research work. He has always motivated me and was so generous with his knowledge and time. The completion of this thesis would not have been possible without his advice, and my time at BC Hydro would not have taught me so much without his guidance. I would also like to thank my colleagues at BC Hydro: Wun Kin Cheng, Alaa Abdalla, Nan Dai and Herbert Louie.

The research could not be finished without support from both BC Hydro and UBC Civil Engineering Department. Dr. Shawwash and Dr. Siu provide me a very good research opportunity and funding which allowed me to concentrate on my research topic without any distractions.

Finally, I would like to thank my parents and my girlfriend who were so supportive and they always gave me their unselfish love and encouragement. Without them, I don’t think I could get through the toughest time when I have first arrived in Canada.
CHAPTER 1 INTRODUCTION

1.1. Background

As the main supplier of electricity in British Columbia and the third largest in Canada, the mission of BC Hydro is "to provide integrated energy solutions to customers in an environmentally and socially responsible manner". This mission has been accomplished in stages of growth in which every component of the system is operated more efficiently.

Figure 1-1: BC Hydro Generation Map
Generally, most of the generated electricity in BC is hydroelectric. As shown in figure 1, the BC Hydro generation system, with 30 hydroelectric power generation facilities and 32 reservoirs in 6 major river basins, 27 watersheds and three thermal generating plants, provides electricity to meet the domestic needs in the province of British Columbia. The total installed generating capacity in B.C. is about 13,300 MW, with a total annual production amounting to more than 60,000 giga-watt-hours. Hydroelectric generation accounts for about 85 percent of the total installed capacity, with the balance being generated by other thermal sources such as oil, natural gas, and other resources. In terms of installed capacity, sales, and customer base, BC Hydro dominates in these areas, as it controls over 82% of the installed capacity, and is responsible for 94% of the electricity sold in the Province. However, some Independent Power Producers (IPPs) and large industries also generate electricity in British Columbia. IPPs either sell their electricity production to BC Hydro, West Kootenay or to the local export market. In addition, there are quite a number of large industries (e.g., ALCAN) who also produce electricity to meet their own needs, and their generators can be in the form of co-generation – as a byproduct of intensive industrial processes such as those that requires pressurized steam. The total IPPs and industrial installed generation capacity accounts for approximately 2300 MW, 80% of which is made by industries. A policy statement that the provincial government issued in 1992 highlighted the IPPs role as electricity suppliers to BC Hydro. The policy encourages IPPs to develop new sources of electric power in province. However, IPP future investments are expected to be limited to other generation technologies (wind, woodwaste, etc) as well as those undeveloped areas with low hydropower potential.

Over 90% of the generating electric in British Columbia, totalling approximately 11,200 MW of electricity comes from BC Hydro’s hydroelectric generation systems. The Williston reservoir on the Peace River (40 billion M³), and the Kinbasket reservoir on the Columbia River (14.8 billion M³) are two reservoirs in the BC Hydro system that provides for multi-year storage. The high capacity of the Williston and the Kinbasket reservoirs enables BC Hydro to plan their operations several years ahead. Thermal generating facilities can supplement the hydroelectric system in years with low water flows and when natural gas prices are low.

Generation scheduling is one of the main activities practiced in every power generation system. The main concern of the BC Hydro system operator involves implementing operations plans that ensures a load-resources balance. On the other hand, since electricity markets have been deregulated and restructured electricity markets in the past few years, the BC Hydro generation system is also planned to for short-term and real time electricity trade operations.
Other functions of the system operator include management of non-power needs, such as balancing power generation requirements with the needs of fish, wildlife, recreation and flood control.

The transmission network connects the generating facilities with the major demand centers in the Province and to the transmissions networks in Alberta and the US. Imbalance between generating resources and demand centers has shaped the development of the BC Hydro transmission network. BC Hydro serves 1.6 million irregularly distributed customers over an area of almost 95 million hectares. To achieve that, about 18,286 kilometres of transmission lines and about 55,254 kilometres of distribution lines have been built over the years. Figure 2 illustrates the BC Hydro main serving areas.
1.2. The Decision Marketing Environment

1.2.1. Introduction to the deregulated electricity marketplace

Traditionally, the electric industry in the United States and Canada was vertically integrated where utilities supplied electric power to the customers at regulated rates. The main concerns of the utilities then were to deliver reliable power supply and to maintain a sustainable rate of growth in power supply to meet the demands. This situation have changed as the electric power industry was deregulated, thereby allowing for electricity to be brought and sold in an open market and for independent power producers to compete with the traditional utilities.

In 1991, California Public Utility Commission (CPUC) decided to restructure the electricity power generation industry. There are several basic underlying concerns that were at work in the background. First, it as realized that as the need for more capacity could be met in a relatively inexpensive way by taking the advantage of excess capacity in areas with excess capacity in another (e.g. the capacity needs could results from unusually high peak demand, or an unexpected plant outage). No centralized market that could facilitate these transactions existed. Second, new economic signals were needed for new generation facilities by the utilities and other investors. Without a doubt, the competitive marketplace could provide such an environment. Third, the electricity market could cause and encourage the utilities to lower the costs of generation and that could eventually benefit the end users. Finally, other sectors of the economy, like trucking and telecommunications, seemed to benefit from deregulation and market forces worked very well in these markets.

The subsequent market dynamics that evolved have, however, posed new challenges. Privatization in the electricity industry has brought the need for the introduction of commercial interfaces between the functions of generation, transmission, distribution and retailing. The behaviour of electricity differs from the one of other commodities, as electric power inventories cannot be established to create pricing securities. For this reason, and to remain competitive to achieve business excellence, companies involved in electricity trade were forced to enhance their decision making framework to guarantee the optimal use of the system resources while addressing the risks involved.

1.2.2. The role and state of B.C electric industry

Until recently, the electric utilities in British Columbia were vertically integrated, with responsibilities for generation, transmission and distribution, and for customer services rested mainly with B.C Hydro. However during the past few years, BC Hydro has realized the
significant benefits from electricity trade through coordination of its system operations with other utilities in Alberta and in the U.S. It is widely held that these coordination activities could optimize the interconnected system resources, increase its security and its reliability. It could provide significant financial benefits to British Columbia and could also improve the security and efficiency of the B.C. electric systems. Specifically, it’s beneficial for B.C hydro, as it owns a predominantly hydroelectric system, where surplus water can be stored in reservoirs during wet years and market trade actions can provide significant additional revenues.

Electricity trade is enhanced by the fact that demand characteristics, type of generation, and climate in different regions vary. For example, during the summer month’s high demands for electricity in Southern California arise from the use of air conditioners, while demand in B.C. is low as no air conditioning is usually needed. This means that “hot” spot market can develop in the months of July and August in California. Another example is the high demand in winter months in Alberta, caused by the use of electricity for heating. This coincides with low demand for electricity in California and to some degree (low demand) in B.C. Electricity trade is also made possible since the BC Hydro transmission network is interconnected with transmission networks in Alberta, West Kootenay Power in south-eastern B.C., the Alcan system in the North Coast, and the interconnected system in the western United States. Currently, the tie line to Alberta provides interchange capacity of up to 1100 MW, while the ties to the U.S. provide interchange capacity of approximately 3250 MW.

Several conditions make electricity trade a viable option for utilities to consider. First, in wet years, surplus water stored in reservoirs can be used to generate and trade electricity in the market. Since BC Hydro possesses considerable storage capability, surplus water can be stored and converted to electric energy. When the storage capability is not sufficient to store the total volume of water inflows, then water is spilled, and the opportunity is lost. Second, in dry years when the water level is low, shortages in energy and capacity can be compensated for by purchases of surplus energy from other utilities. Third, in emergency situations, electricity can be purchased elsewhere to provide for backup power. For instance, if a number of generating units were suddenly taking put out of service, or when a major transmission line is de-energized for a fault, other neighbouring utilities can be called upon to provide support. To prevent blackouts (or brownouts) the procedures to reinstate these system disturbances are usually automated. Fourth, electricity trade can also be used to “time-shift” the generation of other systems that do not have reservoir storage and peaking capability –a key feature of hydroelectric generating facilities.
Fifth, significant revenues can be earned through simultaneous sales to and purchases from other electric utilities (arbitrage).

In one aspect, B.C hydro is dedicated to meet the domestic demand and like other companies, B.C Hydro purchases electricity to maximize the benefits from the marketplaces. This is a balancing act to how effectively and reasonably use the limited water resources in a sustainable way. British Columbia is expecting a growth in demand for electricity of about 3.7% per year for the next few years. To meet this growth, BC Hydro faces a number of tradeoffs between costs, resources availability, reliability and social and environmental impacts. Structural and Non-structural constraints, such as transmission lines total transfer capability and tariffs restrict the way in which electricity can be generated, distributed and traded. Moreover, although hydropower has exceptional qualities among other power generation alternatives, a system with a 90% hydro share carries a large number of uncertainties that need to be properly addressed. All of these concerns drive B.C Hydro to develop an efficient management framework, that could effectively guide system operations in a more optimal and economical way. Indeed, during the last decade B.C Hydro has been developing a number of optimization models that could be used to guide system operation in the most optimal and economic way. It should be noted that those optimization models are used by BC Hydro system operations engineers to system operations.

1.3. Statement of the Problem

Maintenance Scheduling (MS) is a general activity that is widely carried out by all kinds of industries. It is usually carried out to achieve two main objectives: to maximize economic efficiency and to maximize the lifespan and availability of equipments. Generally speaking, the purpose of MS is to maintain the equipments and facilities in a satisfactory operating condition by providing for systematic inspection, detection, and correction of early failures either before they occur or before they develop into major defects. In electricity utilities, the task of maintenance scheduling usually refers to the determination of the outage timings for generating units in service. In this sense, the maintenance of generating units in the electric industry ranks high among the functions performed by the system operators and/or planners. This thesis will focus on developing a modeling methodology for deriving an efficient unit maintenance schedules for hydroelectric systems.

Unit maintenance scheduling is a challenging topic. The system operators usually make decisions to take units off while considering the current units’ working status and the total system operating state. These decisions become more complex as the traditionally vertically
integrated electric utilities are being restructured and deregulated and as new power markets structure and rules become more developed and complex. Currently, market forces, represented by electricity price signals become the main driver in making the unit outage decisions. It is well known that in the electric industry that an appropriate maintenance strategy will have a major impact on a utility’s competitiveness. In addition, thermally dominated utilities also have different concerns than those who are hydro dominated. Thermally dominated plants have a major concern on fuel prices (such as natural gas), unit forced outage rates and unit commitments, since thermal units are usually loaded to satisfy system base load and therefore their commitment will have a major impact on system security or reliability. Hydro plants share most of the same concerns with thermal plants but with added complexity. In particular, and for plants with large generating capacity and with multi-year storage reservoirs, hydrologic water conditions becomes an important factor as they affect the system capacity and the total energy that can be generated over the planning horizon. Moreover, hydro unit maintenance will also have important impacts on the system import/export capabilities in some systems. In particular, system stability and reliability may require that a minimum number of units be online, and in the BC Hydro system this constraint have been called the rotational energy constraint, and it have a major impact on the level of import capability into the system, regardless of the current market conditions. So the essence of unit maintenance scheduling in hydro plants is to develop the unit outage schedules while minimizing the negative impacts on the system reliability and profitability while taking advantage of higher plant efficiency of the hydro system under different unit commitment configurations.

Maintenance scheduling may be carried out with different focus and different corresponding methodologies. For example, reliability-centered methods (RCM) are considered to be the main approach when dealing with the outage scheduling on a component level. It is basically a probabilistic based analysis. In this thesis, maintenance scheduling will be developed at the system level. In another word, the analyses and algorithms developed herein are intended to help the decision-maker to schedule the units’ outages so as to maximize the plants efficiency at different time periods and to maximize the benefits from trading electricity and ancillary services in the electricity market. Several modeling approaches that could potentially be used for maintenance scheduling are reviewed in Chapter 2. Among these, the linear programming method will be investigated in this thesis, particularly those that are based on binary linear programming. Linear programming models have some very obvious advantages, such as the well-developed commercial algorithms, fast solving time and very useful sensitivity analysis.
Furthermore, the Generalized Optimization Model (GOM) was successfully developed for the B.C hydro system and has provided a solid platform for developing the maintenance scheduling optimization model. As it will be outlined in this thesis, the maintenance scheduling model will incorporate most of GOM structure and it will incorporate outage decision activities. In a sense, this thesis can be considered as an extension of ongoing GOM development.

Some characteristics of electricity market are common to those for common commodities markets. But due to the non-storability characteristic of electricity, it is considered to be much more complex than any other markets. In this thesis only deterministic techniques are investigated and they constitute the basis for stochastic analyses that could be carried out in the future. Specifically, the energy demands, water conditions and market prices are assumed to be known beforehand. Although these assumptions will result in incomplete recommendations, the main focus of this thesis is to capture the basic characteristics and properties of unit maintenance scheduling for hydro systems. It will outline the factors that will influence the outage scheduling for hydro units and define how they could be incorporated in an optimization model.

The above description provided a general idea of how the B.C Hydro system is operated. In fact, there are more practical requirements in normal system operations and unit maintenance scheduling is one of them. Reasonable and appropriate unit outage scheduling can not only keep the generators in good working condition, but could also provide a hedge against market risks. Basically, maintenance scheduling could be treated as a sub-topic of the hydro system operation problem, as it “inherits” all the uncertain properties of a conventional hydro system. Moreover, the maintenance scheduling problem could add new “uncertainties” that originate from the units’ outage decisions and those coupled with other uncertainties makes the problem intricate and hardly to solve. The maintenance scheduling algorithm developed in this research will address the most important generating facilities in BC Hydro system. Although the stochastic properties are not the main focus of this research the deterministic algorithms are investigated, which will act as the basis for future development of a stochastic maintenance scheduling models.

Figure 1-3 illustrates the sources of electricity supply in British Columbia in 2004, and it clearly indicates the high percentages of supply from the Peace and the Columbia River systems. It also reveals that the four big plants, GMS, PCN in Peace River and the MCA and REV and ARD on the Columbia River, account for nearly 72% of total system potential capacity. Generally, optimal operation of these five plants can have a major impact on the balance of total BC Hydro system. Hence, this research will focus on developing a maintenance scheduling algorithm and modeling methodology for the generating units of the five plants listed above.
1.4. Goal, Objective and Study Approach

The main goal of this research work is to develop and implement a maintenance scheduling optimization model that serves as a decision support toll for the operation of the BC Hydro's reservoir systems. To achieve this goal, several objectives were identified, first, the BC Hydro system was investigated. Second, a literature review was carried out in hydroelectric system operation and on maintenance scheduling in general and on maintenance scheduling of generating units in particular. Third, formulate the maintenance scheduling problem for a number of hydroelectric generating units. Fourth, investigate and develop the algorithms and techniques that could be used to solve the maintenance scheduling problem.

1.5. Organization of this thesis

Chapter 2 provides a literature review on the maintenance scheduling problems in general with an emphasis on maintenance scheduling in the electric power industry. Chapter 3 outlines a detailed formulation of the maintenance scheduling problem as a linear programming model. Basically, all the aspects affecting the maintenance activities of hydro units are discussed and their relevant formulas are presented. Chapter 4 presents a case study that illustrated the results of the methodologies and algorithms discussed in chapter 3. Finally, Chapter 5 summarizes the main conclusions and outlines prospects for future research and development.
CHAPTER 2 LITERATURE REVIEW

2.1. General Overview

2.1.1. Broad Review of Maintenance Research Development

"Maintenance can be defined as the combination of all technical and associated administrative actions intended to retain an item or system in, or resort it to, a state in which it can perform its required function" (Dekker, 1996). Maintenance activity, particularly in the field of industry, has proven to be extremely important in a highly industrialized society. With more and more advanced and high-tech machines being developed for industry, their safety, productivity, durability and reliability have a significant impact on every-day life of people. Every issue relevant to maintenance, whether from the financial or social point of view, has been the focus of research by many investigators. The importance of maintenance has already been acknowledged to be too overwhelming to be ignored. As an inevitable consequence, system operators attempt to use the latest techniques to precisely describe and measure the behaviours and the corresponding consequences of these maintenance activities. In particular, optimization techniques are usually employed as they help the system operator to understand maintenance performance, and so to make the best “decisions” or evaluations regarding their influences. With this in mind, it is necessary to understand how maintenance studies have developed.

Historically, the first technical approaches aimed at maintenance problems were published between the 1950’s and 1960’s. The aeroplane industry is reportedly the first industry where rigorous maintenance approaches were implemented in order to ensure aircraft safety and reliability. At the initial stages, maintenance studies were mainly reliability-centred (RCM) because their most crucial concerns were the reliability of key operating components. Heretofore, this is still one of the important criteria widely adopted in many industrial areas. During 1970’s, one methodology called ‘conditioning monitoring’ was put forward, focusing on techniques that could predict failure by using the information on the actual state of running equipment. As of the 1980’s, computer models gradually became employed for all kinds of maintenance problems. From the 1950’s to the 1980’s, research studies on maintenance problems had several traits in common. For example, they all attempted to facilitate administrative processes and make management information readily available. In addition, early researchers kept struggling with quantitative methods due to the limitation of computational conditions. After the 1980’s, the main tendency of maintenance research centered on efforts at using computers to simulate large and real systems. Innumerable papers and reports were published during this period, presenting
the great power and potential of using the latest computer and mathematical techniques to handle maintenance problems. An increasing number of practical concerns were involved and the fact that systems were becoming larger and larger meant that the models' complexity had to increase to the extent that it became more and more difficult to control. All these facts indicate that maintenance scheduling is never an easy problem, and that further delicate development must still be conducted in the future.

Regarding the hierarchy of maintenance problems, the most common classification has traditionally been done with respect to the time scale involved (Dekker, 1998). Maintenance activity can be effected by decisions that were made early in the design phase because at that time, the choices, including the type of equipment, the level of system redundancy and the accessibility of machinery fittings would all strongly impact system conditions for the future. However, as opposed to the design phase, most maintenance models focus on the operations phase, especially in inspection or/and preventive models, which primarily target optimal maintenance intervals. To be more specific, during the operations phase, the maintenance optimization model is divided into planning, scheduling and control problems. Another type of common classification involves the levels at which maintenance decisions must be taken. The levels into which the classification can be divided include national or company-wide, plant/system, unit/equipment and component. From a practical viewpoint, this is a very logical way of dividing the classification, and therefore has been widely adopted in industry.

Apparently, maintenance tasks differ very much in terms of their different requirements. In general, several kinds of objectives have been reported in the literature, and they can be summarised into four categories: ensuring system functions (availability, efficiency or product quality), ensuring system life (asset management), ensuring safety and ensuring human well-being. No matter what their objectives, nearly all maintenance optimization models can only place emphasis on one of categories, and try to ensure a balance between that and the remaining ones. Furthermore, despite the variety of models that exist, they still all share some common aspects. For example, there are similarities between the descriptions of different technical systems, including systematic performance or deterioration functions with respect to time, and the descriptions of the available information regarding system operations and actions that are open to the management.
2.1.2. Common Maintenance Models and Their Applicability

Dekker (1996) pointed out that after summarizing the literature, the author found that the most commonly applied model is one done for the purpose of age and block replacement. This type of model has two characteristics. One is that many models have infinite time horizons and the central targets are to measure the optimal average cost for each maintenance interval. The second characteristic is that many of its applications, usually associated with statistics and/or probability analysis, calculate the indices for each component or the whole system in terms of reliability, availability, dependability or applicability. However, these indicators are always widely suspected in view of unrealistic mathematical hypotheses and data collection problems. For example, as stated by Tan (1997), many applications assume that a binary number, which means that only working, presents the component state and malfunctioning statuses can be represented during equipment operations. Furthermore, frequency of failure is the only function that is represented with respect to time, and it has a monotonously increasing tendency. In fact, this is not true in practice. This is because more working statuses of components exist in operations, and there are more factors affecting the deterioration process. Although so, it should be noted that nearly all such methods stem from the framework known as RCM (Reliability Centred Method). RCM is the first failure mode with quantificational analysis. The goal of RCM is to identify the maintenance order, interval and extent of maintenance for each component, taking into account the safety requirements, costs and other factors.

The other model often used is known as the Markov Decision Model. The Markov decision process is usually depicted as a graph consisting of nodes (states) and arcs, and is suitable for situations that require component replacements with multiple states and transition probabilities. Therefore, system stochastic properties can be handled very intuitively using the Markov decision process. But cases are sometimes too complicated to apply Markovian approaches, for example a few reported applications belong to semi or hidden Markov processes. Some simplifications or transformations must be done such that the applications could be solved using a traditional Markovian approach. Moreover, for situations that use multiple interacting components with time-dependent failure frequencies, the Markov model no longer seems to be practically applicable. This is because when state spaces (usually because of multiple components) increase into multi-dimensions, solving the problem using Markov decision models becomes an impractical task due to the extreme need for large amounts of computer memory. Hence, one common way to deal with the situation has been to relax certain assumptions or divide the original integral model into several separate ones. This approach renders the problem
of scale more controllable. Even though there are many theoretical or practical restrictions for both models, many successful situations such as equipment overhauls, vehicle replacement, constant appearance of new devices, road maintenance and scheduling of overhauls of electricity power stations are still based on them.

In addition to these traditional maintenance models and their applications, other newer models have also appeared in the fields of maintenance. As indicated by Scarf (1999), “the most recent research has been carried on condition-based maintenance, maintenance of multi-component systems and inspection maintenance”. First of all, the so-called condition-based maintenance model suggests that component or system availability may be determined more accurately and meaningfully by ways other than through the use of traditional probability analyses. For instance, by using certain engineering approaches, the descriptions used for various equipment conditions can make more authentic sense than those derived from statistical estimates. But the difficulties inherent in realizing condition-based models may occur in following ways: 1). insufficient knowledge in determining the operating status of each piece of equipment; 2) the challenges of integrating this knowledge into an appropriate decision model which must also consider maintenance costs.

Furthermore, although there is a very long history in this field, the research on multiple component maintenance is still restricted to very specific cases, and mostly consists only of pure mathematical investigations. The main reason for this is due to the difficulty in obtaining meaningful statistical data.

In addition, the inspection model, also called the delay-time model, has only very recently been developed. It concentrates on the frequency of inspections during which the pre-stages of failure, or what are known as “faults”, can be observed. Hence, the delay-time is the time that elapses between the first moment when a fault can be observed and the eventual failure. But two facts make the model very difficult to generalize from. One is that delay-time seems to be age-dependent in almost all industrial studies, and the other is that minimization of cost is usually not the main deriving force, as has been seen in the case of medical screening studies.

Finally, there is the capital replacement model. This model has become a part of short-term maintenance modelling because strategic planners have started to show concern regarding the provision of resources to guarantee future competitiveness under various kinds of economic pressures. A new discipline that involves teaching maintenance modelling has even appeared. This discipline focuses on working with equipment users and informing them on the theories of maintenance by using simple and understandable case studies.
As we have seen, the main reason why maintenance models can't be easily organized is because the statistical data associated with component and equipment failure is very hard to properly collect. Thus, many suggestive approaches, including gathering information from experienced workers, have been proposed. Nevertheless, several of those approaches are case-specific. As an exception, one alternative model that is seemingly general, involves the Operational Decision Support System (DSS) or Maintenance Management Information (MMI). Although they are still far from being the ultimate solutions, they do provide innovative potential to integrate various historical data, expert opinions, logistical reports, plant functions and other useful resources. Moreover, with general operational decision support systems, people can easily answer the "what-if" questions that provide answers and choices in comparing different maintenance alternatives.

2.1.3. Generalized Maintenance Frameworks

Although one can say that every maintenance optimization model is very dependent on its particular background, there are still some researchers investigating the potentialities of a generalized framework. Two papers are worth mentioning here. One paper is from Vatn et al (1996), who realized that RCM does not originate from an optimization angle, so an optimal strategy for maintenance could not really be derived from it. The authors attempted to extend the traditional RCM into a wider range by including both system (and/or component) cost loss functions and reliability (dependability) indices together. Through the use of linear programming or heuristic searching, the system costs loss and reliability requirements are weighted and balanced so as to determine optimal maintenance intervals. Besides, through their case study, the authors provided an example to show the procedures through which to analyze and completely define the target system. Finally, the model let decision-makers realize the importance of incorporating possible failures into the decision-making process. Although this paper primarily involves a trial, it does present the insight and perspective of one general idea.

Dekker (1995) also addressed an integral model that considered optimization, priority setting, planning and combining maintenance activities. In this paper, the author provided an analytic method of considering priority setting, planning and combining maintenance activities. The central contribution of this paper involves a new concept known as "marginal expected cost", which refers to the outcome of deferring the execution of maintenance activities for an infinitesimally small interval. A theorem is first developed to depict the behaviour of the deterioration rate and to prove the existence of minimum average cost. Aiming at different types
of maintenance systems, specific deterioration functions are then possible to establish. From this, one can predict whether there is a duration when the costs caused by deterioration could be minimized. Models are also extended for various situations and considerations. After determining the optimum maintenance interval, penalty cost function is applied in order to measure the costs of deferring or advancing this outage activity. Then, based on the minimization of total penalty costs and other experience criteria, combinational maintenance cases could be readily modelled and determined. The only deficiency of this paper is that it fails to show much applicability in real surroundings. But, as an initial development of a general framework for modelling maintenance scheduling, its meaning is obviously more significant than its deficiency.

Finally, it is worth pointing out that the main industries where maintenance systems are of great concern and highly researched are the oil & chemical industries, as well as industries concerning railways, transport companies, airlines, steelworks, electricity utilities, defence and civil infrastructures.

### 2.2. Maintenance Scheduling in Electricity Utilities

#### 2.2.1. General Introductions

The topic of maintenance scheduling problems in the electrical power industry is not a new one, and early research work in this field can be traced back to the early 1930s. The literature indicates that most of the maintenance scheduling studies were carried out at the unit/plant level. The main objective of analyzing or optimizing maintenance activities was to maximize economic benefits and system reliability. Compared with other industries, maintenance scheduling in the electrical power industry has specific properties which include:

1) The scale of the system is usually very large. Unlike maintenance situations in other industries, the decision-makers in an electrical power system are always facing the problem of needing to maintain tens, or even hundreds of units. Usually, the solutions offered by the optimization model for such a system leads to a very challenging combinatorial problem. As a result, many researchers have focused on finding the best solution to such problems using computational approaches showing a high level of efficiency.

2) The uncertainties associated with maintenance problems in the electrical power industry are relatively easy to obtain and assess, and therefore statistical data from these problems can be easily derived and analyzed. With more convincing data, reliability analysis can be readily carried out by using probability theories.
3) Although reliability is still one of the greatest concerns for maintenance scheduling, more and more people have started realizing that minimizing costs will play an increasingly important role in the electricity industry. Additionally, there has been relatively less research reported on maintenance models at the component level - for electricity utilities the main reliability analysis is still based on the total system.

In his review paper, Yamayee (1982) summarized the basic perspectives of past research interests. In addition, he has clearly identified and defined important terminologies that have appeared in past maintenance research. He also identified and discussed different types of costs and reliability objectives that he found in papers that were published prior to the 1980’s. Basically, the authors identified that costs should be divided into maintenance costs and production costs, but they did not believe that minimizing production costs would be effective because production costs seemed to have quite insignificant impacts. Reliability analysis can be categorized into deterministic and the stochastic cases. Deterministic reliability is based on the concept of levelized capacity reserve, while stochastic reliability is mainly attributed to the randomness in the load forecast and to the generators’ Forced Outage Rate (FOR). Regardless of the type of reliability considered, it seems that the issue of reliability has been the main focus of past research activities. The author provided an interesting perspective that suggested that the optimal maintenance scheduling problem should be integrated into total system operation models at various time scales. For example, in a hydroelectric system the long-term models would optimize the multi-annual storage in reservoirs; seasonal models would provide the best annual maintenance scheduling and operation strategy; and finally in weekly or even shorter time scale models, economic dispatch, unit commitment and short-term storage become the main concerns.

EL-Sharkh et al (1998) presented relatively recent developments in maintenance scheduling studies. The main discussions still focused on cost and reliability, but the cost objectives were no longer believed to be possible to ignore or insignificant. On the contrary, the paper highlighted and stressed their importance, particularly in a deregulated market environment. Moreover, the costs of reserve were introduced to reflect the costs representing reduced reserve margins caused by generating unit outages. On the other hand, unlike in some of the past literature, no significant changes were suggested for reliability analysis. However, some new constraints and indicators were introduced, including as the Loss of Load Probability (LOLP) or Expected Unsupplied Energy (EUE). The paper also summarized general maintenance constraints. Basically, these constraints can be divided into the requirements of maintenance, isolated systems and interconnected systems, as well as reliability limitations. The
following is a list for general consideration of maintenance scheduling in the electrical power industry:

- availability of maintenance crew constraints;
- resource/equipment availability constraints;
- time/season constraints;
- system/plant maximum allowed outage capacities;
- unit capacity constraints for both thermal and hydro units;
- supply/demand constraints;
- transmission constraints;
- power sales or purchase constraints;
- member scheduling for electric power pool coordination constraints;
- Special thermal constraints such as coal supply, transportation and production.
- Special hydro constraints, like reservoir storage, spill or fishery constraints.
- Reliability constraints, like LOLP, EUE and the expected energy generation (EEG) constraint.

Although both Yamayee and El-Sharkh talked about methodologies applied to maintenance applications such as linear and dynamic programming, those discussions aren’t detailed enough to understand the whole scope of the techniques and their particular details or characters. Therefore, reported applications and papers are categorized by their applied methodologies. Meanwhile, in order to help people clearly grasp the main ideas regarding the development of these approaches the papers by different authors are ordered in chronological sequence. The papers of authors have been collected together, and arranged in chronological sequence. Thus, it allows people to develop a clearer understanding of how and why specific methods were developed. Finally, it should be noted that many of these papers appeared in IEEE Transaction on Power System.

2.2.2. Linear / Integer Based Programming

Dopazo (1975) presented a well-known paper regarding the integer (0-1) model for maintenance schedules. The model focuses mainly on cost functions, and does not concern itself with risk levelization. The author does, however mention that his model could transform reserve considerations into constraints. Several optimization techniques are discussed in order to solve this 0-1 problem. From a current technical standpoint, this paper is not particularly advanced, but
it was very significant at the time of publication because it first introduced linear programming, which showed a powerful potential for scheduling the unit outage.

Probably because of difficult-to-solve inherences in integer programming models (IP), as well as the low computational capabilities that existed in the past, integer models were seldom reported before the 1990’s. Mukerji (1991) focused on a practical example and built an integer programming model to balance the different objectives of economy and reliability so as to get a practical outage schedule. Two special points were mentioned in this paper; first, the author used a cumulant approach to calculate production costs because he thought production costs were convoluted with system loads and the units’ efficiency and availability, which were believed to be difficult in the past because of their high non-linearity and random natures. Secondly, based on two criteria/objectives, i.e., reliability and cost, he ran the optimization model several times to investigate the mechanism of how costs and reliability would affect each other. In addition, according to the different purposes, he also researched the impacts of the relevant constraints by relaxing them somewhat. Results showed that the two criteria would have different optimal schedules and in real situations, they should be coupled together.

Chattopadhyay (1998) proposed an IP model in consideration of fuel/coal supply and transmission. Herein, the binary variables were designed for the purpose of rendering the maintenance duration continuous. The author realized that the IP problem is not easily solvable, so for the first run, he relaxed the integrality requirement that made the IP problem degrade into a linear model. Then, by virtue of some expert rules, he analyzed those “split” integer variables and adjusted them to be integers that represented the consecutive durations for plant maintenance. Although this method deviated from the real optimal value, the faster speed and practical results were still deeply impressive. After fixing all consecutive outage periods, the author added peak demand constraint to the third step, which had been omitted in the first step because this constraint couldn’t be expressed without using binary variables. Furthermore, the author introduced Monte Carlo simulation into this linear model in order to capture the random availability of generators, also known as forced outage rates. The objective function in this paper was tested respectively as total cost, variance of EUE (Expected Unserved Energy) and total EUE. The case study showed that the objective of total cost had higher preference. The inadequacies of this paper were felt to exist in three places. First, the model included the constraints describing the supply/transportation requirements of oil or coal. Those constraints no longer makes sense because the transportation itself is a quite a complex problem and needs to be further researched. Second, the behaviours of random variables such as availability of generator
and mine price fluctuations were not sufficiently investigated; so using Monte Carlo simulation was very sceptical. Third, the term hydro constraint is too simple to be meaningful.

Chattopadhyay (2004) extended his model to life-cycle span that still kept use of the Monte Carlo techniques for simulating random Forced Outage Rates (FOR). To do this, he first built a base linear optimization model that maximized the Earnings-Before-Interests-and-Tax (EBIT). A noticeable point of this model was the new presence of economic aspects, such as forward contracts and pool revenue, as well as dynamic unit performances due to their (non) maintenances. After solving the base optimization model and fixing/dropping some of variables, Monte Carlo simulation could determine a range of EBIT if different FOR samples were taken. Similar to his previous paper (1998), it was found that single random variable (FOR) was not quite enough in a real competitive environment. An increasing number of uncertainties would certainly be preferable. Moreover, forward contracts and spot prices for pools were too complex to be predictable in the long run, especially in a life-cycle span (22 years used in his case study), and consequently the results seemed less convincing.

Leou’s paper (2001) dealt with fuzzy integer programming. The idea of this model originated from the situation that with the objective of minimizing cost functions, including spinning reserve constraints would usually result in conducting “earliest possible” maintenance. Furthermore, this spinning reserve constraint was reported to be the possible reason for the lack of feasibility of “crisp” integer programming. Facing this problem, the author realized that if the model would allow some violations against the constraints, and if this method could produce one feasible solution, the final schedule was still considered acceptable for the decision makers. So, based on this reasoning, fuzzy membership functions that could allow certain constraint violations in the original 0-1 IP model were created. Leou believed that the case/run with the least violations (or the largest satisfaction degrees of constraints) would be picked for the final and optimal scheduling. The author also presented a method, called backtracking skill, to find a solution for fuzzy 0-1 integer programming.

2.2.3. Reliability Centred Analysis

Former studies on maintenance problems were largely based on traditional reliability theories. The reasons for this are twofold. First, computer capabilities in those days were not widely accessible and powerful enough, so only analytic frameworks were concentrated on in maintenance research. Under such circumstances therefore, reliability-centred approaches were the most frequent focus of investigation and were widely accepted. On the other hand, thermal
units were usually the main systems of concern because 1), most large utilities were based on thermal units; 2), thermal units usually had mass outputs and consequently had serious impacts on system reliability and safety; 3) in terms of operational requirements, it was not allowable to turn down and start up thermal units on a frequent basis. As a result, the decision to take thermal units into maintenance wasn’t a simple one. The preceding characteristics resulted in them always undertaking the base load in the system. So, in comparison with hydro units, the operators of thermal systems were more concerned about the units’ stability and reliability. In the early periods, reliability was the only indicator for evaluating a given maintenance plan, with no focus of consideration on “optimality”. Gradually, by introducing the concept of optimality, reliability began to evolve into optimal criteria or objectives in the modelling. Since then, the “integration” of traditional reliability index and optimization techniques have been broadly accepted and generalized within the electrical utilities industry.

With this in mind, some typical papers related to reliability methodology in the electricity industry will be discussed in a chronological sequence. One can observe the reasons and processes for their development. Also, reliability analysis based on component levels (essentially RCM methodologies), were also generally accepted and applied in the electricity industry. They will also be introduced briefly at the end of this section, but in view of the fact that this thesis is approached more from a systematic angle, they will not be highly emphasized.

Originating from the fact that the schedules that didn’t have level-reserves may have lower risk associated with them, Garver (1972) published a now-famous paper to illustrate a method of using effective capability to replace traditional reserve. Garver contributed two concepts with this paper. One was that of the effective capacity of a unit, which implied that taking one generating unit out would be equivalent to increasing the load correspondingly in terms of system reliability. Additionally, equivalent load was defined in such a way that if it was encountered the same number of times as varying daily peak loads; it would cause the same level of risk for the unit maintenance interval. It was similar to a “mean” load over the maintenance interval, but in the reliability sense, not by arithmetical “average”. By rescheduling the unit outage with the equivalent load and effective capacities, it was usually associated with a lower risk level compared with the levelling reserve. One obvious disadvantage may exist in this method: the problem of data preparation. People must accurately acknowledge the forced outage rate for each unit so as to calculate the equivalent load and effective capacity. In addition, as essentially a probability approach, the probability distribution of peak loads was needed. As discussed before, this typical reliability approach was not relevant from the point of view of
optimality, and thus was unable to determine the “optimal” schedules for maintenance. This method could not take economic factors into account either.

Christiaanse (1972) first introduced the optimization idea into reliability maintenance scheduling problems. He realized that a practical outage schedule should contain reserve and capital considerations, as well as all types of other actual constraints. However, in his opinion, the variations of total capital costs by different maintenance schedules were small. In addition, the alternative of equalizing the probability of not meeting peak loads (loss of load probability) was comparatively complex and not convincing due to the inadequacy of evaluating unit forced outage rates. As a result, the author suggested a max-min model of maximizing the minimum net reserve. In order to search for optimal solutions, a method similar to a decision-tree or branch-bound was specifically designed. However, from the descriptions about the performance of this method, it seemed to be time-consuming and was not able to optimality guarantee the solutions either. So, a technique was developed which was essentially based on “crew priority” constraint, and was further investigated so as to accelerate its searching efficiency. Since this technique was rooted in practical experience, it was hard to treat with a mathematical approach.

Patton (1972) introduced an approximate approach that was named the “annual risk minimization” method. As the same as reserve-levelized or risk-levelized method, this approach also needed a unit ordering which specified the priority sequence showing the importance of every unit that was going to be maintained. Differently, the algorithm proposed in the paper suggested that the outage scheduling should be “optimized” under the objective of minimizing annual risks, but only in consideration of the currently chosen unit. After locating the outage duration that made the system reach a minimum annual risk, one can fix this schedule and do the same thing for the next one. It was clear that this method was “sub-optimal”, because the author assumed that the non-linearity of reliability function induced by different grouping combinations of units could be evaluated or represented by linear summations of the reliability index of each unit. By comparing three scheduling algorithms, the author concluded that none of those three scheduling algorithm always gave the results as measured by the resulting annual risk. Furthermore, none of them had a significant advantage over any other method.

Zurn (1977) put forward four main considerations for scheduling unit preventive maintenance in his paper. These included operating cost, reliability, deviations from a desired maintenance scheduling and constraint violations. In this paper, the first two aspects were thoroughly discussed. For thermal systems, the production cost and reliability index, such as Expected Duration of Unmet Demand (EDUD) or Loss of Load Probability (LOLP) could all be
determined by the stochastic model of Baleriaux and Booth. Zurn postulated the following: since the random unmet demand (or reserve) was equal to the difference of the random demand and random generation, the unmet demand distribution was equal to the deconvolution between the demand distribution in a standard/representative day and the density function of generation availability. This was obtained by successive additive convolutions of the individual unit availability density functions. If a unit was removed from service during a given period, the load it would have carried would then likely have to be undertaken by more expensive generation according to the pre-established unit commitment and loading order. The difference in energy production cost is thus obtainable. The case study showed that among all reliability indices, EDUD or LOLP got the best results.

Stremel (1980) developed Booth’s method into a very easily computable model by using cumulant and Gram-Charlier expansion to represent the equivalent load (the convolution between load and outage) distribution. The advantage of his new model manifested in faster computational speed and easier implementation. The “expected” generation of a unit could readily be determined by algebraic expressions. Based on the same mathematical formulas, Stremel (1981) applied this cumulant method to the process of evaluating the levelling reserve in maintenance scheduling cases. Compared to Garver’s method, the new model seemed to attain the results with less system risks. The reason could be explained as follows: cumulant method considered varying load-carrying capacities of each unit as they were removed from the system in different ways, instead of from the full and static system capacity outage table used in Garver’s method. As pointed by Stremel, Garver’s approach was only reasonable for the first unit in the maintenance period, but may be quite inaccurate for any other additional units.

Billinton (1983) et al. adopted the same fast cumulant algorithm to compute equivalent load. They presented two case studies and contributed two main conclusions in their paper. First, by comparing the traditional levelizing reserve with the levelizing risk method, they concluded that when using a risk-levelizing approach, the reserve was larger but LOLP was lower than the corresponding quantities obtained by the reserve-levelizing method. Second, load and peak forecasting and unit FOR were sensitive and could seriously affect the maintenance scheduling.

Yamayee (1983) used the fast cumulant method to compute production costs in an optimal model. This model had two objective functions - one was production cost and the other was reliability index (LOLP). Dynamic Programming Successive Approximation (DPSA) was implemented to solve the model. The case study arrived at two conclusions. One was that the difference of production costs between two objective functions was quite close. The other was...
that the index of LOLP and running efficiency in a reliability-objective case was much better than that from a cost-objective case. Although the author mentioned that the actual maintenance scheduling for the two objective cases were very different, he didn’t specify and compare their strengths and weaknesses. As a result, it was unclear whether or not the reliability objective was preferable. Meanwhile, since DPSA was reportedly a sub-optimal method, the conclusion drawn by DPSA seemed somehow suspicious.

Contaxis (1989) carefully presented the procedures for evaluating the LOLP using an analytical approach and various numerical examples. He discussed and compared three maintenance scheduling algorithms sequentially: levelizing reserve algorithm, levelizing effective reserve algorithm and levelizing incremental risk algorithm. The third one was a quite novel reliability indicator that reflected the idea that people who wanted the system would be minimally impacted by the units’ maintenance. It could also be understood such that the risk changes by taking a unit out would be as low as possible in the whole study period. The conclusion proved that this method was indeed the most effective one. Additionally, all three methods were heuristic-based because they needed certain priority lists.

Juan and Ortega (1997) conducted challenging investigations on rational considerations regarding hydro units when calculating the system reliability indicator LOLE (Loss of Load Expectation) and EENS (Expected Value of Energy Not Served) in the models. Traditionally, a hydro units’ outage rate was ignored and taken as zero because first of all, in many systems, hydro capacity was small enough to be ignored, and secondly because they could be turned on and off easily and this impacted system reliability less. Hydro units can’t be applied by cumulant approximation as thermal units can. The conventional way would be to make sure that the area under ELDC (Effective Load Duration Curve) between only thermal capacity and thermal plus hydro capacity should be equal to, or less than AHE (Available Hydro Energy). But this method usually overestimates the reliability of a generation system. So, the authors proposed an approach that “discounted” hydro energy and “subtracted” it from the ELDC. Then, the “remaining” curve was only valid for thermals. This procedure facilitated the maintenance ordering of thermal units in terms of either minimization or levelization of the LOLE index, because after deconvolution of the hydro energy, the reliability indices were counted by thermal units. This ordering was performed by heuristic search and priority setting and not by an optimization algorithm.

For a larger hydro system, obviously hydro resources lead to the improvement of system reliability, but they also caused “unstable” impacts on reliability evaluations because of seasonal
water inflows. Therefore reasonable allocation of the water energy becomes very crucial in that the system reliability indicator (LOLE) would be preferable if all time intervals (month, week or days) could have the same reliability values. Gonzalez and Juan (1999) proposed an approach to solve this problem. The basic idea was taken from their previous paper. But the difficulty now was how to guarantee that the summation of the allocated water energy in each period/interval would be equal to the annual total water energy, under the prerequisite of keeping the same LOLE for the whole duration. The authors presented their steps to solve the problem. Although the hydro maintenance of units was not the main concern of this paper, the main ideas from this paper can be borrowed and applied to maintenance problems.

In a deregulated electricity marketplace, hydro systems manifested more and more of their exclusive advantages: hydro capacity can be sold into a reserve market where the prices are much higher than spot prices. Consequently, people involved with thermal systems considered the problem of how to schedule and allocate the water energy to market in a manner that could not only bring in maximum profits but also keep system reliability during the outage seasons of thermal units. Camino and Juan et al (2005) continued working on this problem. Their approach, still originating from the “modified” ELDC idea, realized that under the constant reliability requirements, water capacity could be maximally sold to the reserve market. But still, hydro units were not responsible for system reliability, which was only provided by thermal units.

Differing from some of the mainstream papers, some scholars, such as Vaurio (1995) put emphasis on the risks of separate components and their corresponding costs to the maintenance activity on the plant level. This was an analytic paper that is based on the minimization of the total plant-level cost under the constraints that the total accident frequency (risk) must remain below a set criterion. All of the risks were measured by the time-average values of accident rates. The whole procedure was formulated in terms of minimum cut terms (sets). Although the work itself was quite insightful and inspiring, the systematic assumptions seemed too “idealistic” to be of practical value in the real environment.

Bertling et al. (2005) proposed a method for comparing the effects of different maintenance strategies on system reliability and cost. This method related reliability theory to the experiences gained from statistics and the practical knowledge of component failure and maintenance measures. In particular, a function relationship between failure rate and maintenance measures had been established for a cable component. As with the papers discussed in the general review section, this paper was quite typical of those of the RCM (Reliability Centred Method) based approach. This kind of methodology required too specific component
"failure" data to be suitable for system level maintenance. The lack of suitability also resulted from the concern that the costs evaluated in RCM didn’t contain the opportunity profits of gain and loss in competitive surroundings, which were becoming more and more emphasized by many electricity utilities.

Endrenyi et al’s paper (1998) could also be classified into the RCM category, and is, based on component-level reliability analysis. The central contribution of this paper was first of all a probabilistic modelling composed of component ageing; and secondly, the balancing of the opposite effects by component failures and maintenance outages (or costs), and thirdly the prediction of remaining life given that a certain stage in the ageing process had already been reached. This prediction was evaluated by a Markov Chain because the component statuses were depicted by state transition probabilities. Their proposed approach was very useful because it could provide quantitative analysis on the reliability indicators under a given maintenance policy, which was usually impossible to do through RCM. The paper concluded that the trade-off between reliability and costs were described so as to identify the optimal maintenance policies. This method has been reportedly adopted by the Ontario hydro system.

Another paper dealing with the maintenance model of Ontario Hydro comes from Anders and Hamoud (2003) et al. However, this paper didn’t fully stand on the methods based on reliability methodology. The reason for categorizing it into the reliability category is that, first, it involved very specific DC load flow calculations which determined system reliability indices, and then, it could serve as a good contrast to the preceding paper by Endrenyi, who also presented a component maintenance strategy for Ontario Hydro. The model presented in the paper provided several meaningful and useful references to us because it seemed quite successful with regard to a real bulk electricity utility. One of their conclusions is that generating unit maintenance should and could be naturally decoupled with transmission maintenance. Another one is that ARC (Acceptance Reference Cost), which was defined as a criterion with minimum incremental cost on the maximum number of components maintenance, was an idea that was very significant from the perspective of system operators. The whole model was implemented as such a sequence: first of all, operators used heuristic searching algorithms to arrive at an acceptable generating unit outage planning, say, in a one year period. Then Monte Carlo Simulation would take care of the evaluation of reliability indices in chronological sequence, in which the transmission model was divided into three regions. Following that, ACR could be calculated and checked whether it was below the predefined level. And finally, a reliability program would again be called to consider a certain maintenance period in a cumulative manner.
If certain indices showed some violations in terms of this outage strategy, adjustment of unit maintenance would be remade, and the whole procedure would be re-run from the top until all the conditions were satisfied.

The census-report paper by the IEEE/PES Task Force Subcommittee presented implementation situations of RCM-based methods by electricity utilities which were based primarily in North America. This paper reviewed the basic concepts of various maintenance approaches and the mathematical models relating reliability and maintenance. Although this report recommended that probability-based models should be developed into policies of utility maintenance, the model complexity, as well as the difficulty of collecting rational data, made those approaches too ideal to be realistic. At least from this report, it was rare for electricity utilities to adopt them as strategy determination tools. Moreover, empirical forms of predictive maintenance based on periodic inspections were still the main tendency for most electricity utilities. Finally, the paper also admitted that in deregulated environments, the deriving force on the decision of maintenance came more from profit incentives than from optimal costs on maintenance and repairs. Maintenance policies in competitive environments could have a little deviation from the initial concerns of traditional RCM.

2.2.4. Dynamic programming model

Dynamic programming (DP) was ever reported in early period aiming to solve maintenance scheduling problem, such as the literatures by Zurn (1975) and Yamayee (1983). DP is known as one of the most proper methods for sequential decision problems. In his paper Yamayee (1983) claimed that DP was the appropriate approach for solving MS problems, due to its characteristic of sequential decisions. Additionally in DP method, the objective function didn’t have to be a continuous function with respect to the decision and state variables, and even both objective and constraint expressions didn’t require the analytical forms. The main disadvantage of the DP or any similar enumerating techniques, such as decision trees, was that it usually required a large number of temporary variables to be saved at each stage. The situation is even worse if the system has two or a higher dimension, which is well known as “curse of dimensionality”. Even using the current computers, the curse of dimensionality is still a problem. Recently, the DP method seemed to be used less than in previously. Except the inherent disadvantages of DP motioned above, some other reasons may be as follows:

- Reliability analysis was not appropriately coupled with DP procedure because the reliability indices were not naturally in connection to system state variables.
In most instances, system state couldn’t be readily defined because the maintenance continuity requirement makes this definition really uneasy. In MS case, one maintenance decision not only affects current state and stage, but also directly determines the decisions for next several stages and states. So it couldn’t be treated simply as conventional “action”. Furthermore, if multiple units’ outages existed, conventional decision variables would become helpless to represent the combinations of possible maintained units.

But the good properties of DP are also very unique. First DP can easily handle the integer cases; and then probability transitions between the states can very intuitive be described. In personal opinion, DP may not be a technique directly and fully suitable for MS while it will be quite promising if it can combine with other modelling techniques, like integer programming to construct a comprehensive model. For example, firstly one could give an assumed maintenance schedule, and then use DP as a platform to handle system uncertainties associated with complex constraints. After that, the outcomes could be input into integer model. Integer model could easily reschedule the combinations of units’ maintenance. If there occur some inconsistencies, correct or relax certain inputs and constraints, back those corrections to DP model and iteratively run two models until all errors are within the satisfactions. Moreover, with the new techniques like parallel computations, the curse of dimensionality is also possibly conquerable. From those angles, DP is till attractive enough to be applicable in maintenance case.

2.2.5. Decomposition / Hierarchical Model

Maintenance optimal problem is hardly solvable, mainly because of those integral and non-linear properties. Many people kept trying to decouple or decompose the original integrated model into several smaller ones according to their functionalities, purposes and other characteristics. The solution by each “separate” model would be adjusted or improved in terms of the special mechanisms such that balances among all the models could be achieved. Then it is possible integrate one acceptable solution which would be treated as an approximation to the original model. Each individual model could be linear or non-linear optimization model. Even in more general sense, it could just be a common procedure with certain purpose, such as the computation of system reliability. Commonly, there would be a “master” model connecting the behaviours or actions among all of its “slave” models. This master one usually presents the main objective of the original model and the salve ones are usually separated in terms of those hardly solvable parts in original model or the constraint cuts that “revise” the relaxation effect by original model. Following are some successful applications ever published in history.
Chen and Toyoda (1990) built a model with hierarchical structure and its main purpose was to search the optimal LOLP. The main point of this paper was that the authors believed that equalizing the increment of LOLP could result in the better maintenance scheduling than equalizing the risk or reserve. They created a concept named virtual load and defined it as a modified load pattern according to the reliability index or reserve margins. Then they constructed a two-level structure and made efforts to solve the complex combinatorial maintenance problem. The system was divided into two models: master problem and multiple sub-problems. Two parts were connected by 'virtual load' that would be dynamically adjusted by master problem for reaching the total minimum LOLP, under the condition that the sum of virtual loads could not be unchanged. The sub-problems were grouped by the units with the same maintenance duration, while master problem used incremental LOLP as a guideline to adjust those virtual loads. This method was quite a new attempt in LOLP algorithm, but in view of "virtual" assumptions, the practical meaning was still somehow questionable.

In succession, considering the relations among interconnected areas, Chen and Toyoda (1991) gave an algorithm to determine the maintenance scheduling. Since conventional methods always ignored the transmission limits or net constraints, it was very essential to have ways taking care of those practical considerations. This paper designed two solving steps. First, sub-problems are responsible for each isolated region and get the maintenance schedule for each unit. Second, master model could take those outage schedules as known information and add transmission limits subsequently. The regional reserve would be minimized as objective function. After that, the so-called "virtual load" for each region was calculated by adjusting regional reserve slightly and then, it would be passed into the sub-problems. Two problems are alternately solved until the convergence criterion was reached. However, there were two noticeable implications. One was hidden in the sub-problem, which stated that the outage schedule was determined by unit ordering or priority. The other was that active flow on transmission lines was controllable. Both sub- and master problem were modelled in linear programming.

Al-Khamis et al (1992) established a model composed of two or three parts according to their different properties. Master problem was a linear/integer programming model aiming at the determination of one potential schedule. One sub-problem was regarding reliability concern, which required that the schedule generated by master problem would cause no violation against reliability constraints. But due to the non-linearity of the reliability index ENS (energy not supplied), Lagrange multiplier method was applied to minimize the acceptable reliability level. If
the current maintenance schedule was not satisfactory enough, the reliability constraint, associated with the Lagrange multiplier, would be taken as an additional ‘cut’ added to master problem, until the solution from the next run was feasible for both of them. Similarly, the same mechanism can also be applicable to the pattern that load must be undertaken by all units (which affect the ENS computation) but simultaneously hold the constraint that each unit couldn’t over-consume the pre-dispatched fuel.

Moro (1999) also used multi-objective model to schedule the units’ outage, which was named in the article as goal programming. From the similar considerations, minimum cost and superior reliability were two main objectives and would be implemented in two consecutive stages. The basic idea was stated as follows: at the first stage the economic optimization model was going to be performed to get an optimal “dollar” number. Then one relaxed this value manually a little bit and took it as a constraint or “cut” into the second stage where reliability criterion was centrally concentrated. The essence idea was very manifest: two objectives must be “compromised” to certain extents so as to get the “balance”. Practically, both of two stages were modelled as integer programming. But the article suggested that hydro units could be aggregated without any special reservoir considerations. For simplicity reason it was doable, but for hydro-dominant utilities, it would be loss of reasonability.

Silva’s early paper (1995) introduced Bender Approach to decompose a very complex, high-dimensional MS problem into two easy-solving master and sub-problems. Their connections were based on so-called Bender Cuts, which was described essentially as newly added constraints in master model coming from certain violations by sub-problem. Mater problem was of deterministic sense, pointing to the best unit outage timing by integer programming. Sub-problem was more concerned about operational conditions, which was often associated with stochastic concerns. However, in the paper, the only uncertain property was from the loads dynamics. Two models were solved iteratively until specified convergence criterion was reached. Although the authors stated that their model also involved hydro considerations, the formulas failed to present this point. Possibly the hydro considerations were oversimplified that they seemed unrealistic in practice indeed.

Subsequent to their preceding paper, Silva (2000) again gave from Bender decomposition angle an integer programming model including both operational (economy) and managerial (reliability) concerns. Naturally, the original objectives were composed of both maintenance costs (operational part) and the costs due to the load cuts caused by outage (managerial part). Its underlying reason was that, the author realized in some particular systems, doing equipments
maintenance may significantly impact transactions accorded previously between energy producer and consumers. And from grid security viewpoint, the unit outage would induce some contingencies, such as the issues of controlling the overloads and supporting the voltage. As a result, a DC optimal power flow formulation was developed to deal with those issues. By decomposing the objectives into two separate costs and solving them recursively, the total optimal cost could be converged. However, the author mentioned that hydro-dominant system could “naturally” be decoupled between production costing and maintenance scheduling problem. This conclusion was quite sceptical because the opportunity cost of large hydro plant was very relevant to time intervals. From the personal viewpoint, it wasn’t appropriate to decompose those two parts in a hydro-dominant system.

Marwali and Shahidehpour (1999) researched systematically Benders decomposition method with network constraints. They firstly built a master problem with integer programming for searching optimal maintenance timing, and then they fixed and input this schedule, as well as network constraints into the sub-problems, and solved them subsequently. The interactive connections between master and sub problems were so-called Bender cuts, which could either be cost cuts, increasing the cost bounds for next master problem iteration if sub-problems were all feasible in current run, or reliability cuts forcing master problem to find other maintenance timing in order to satisfy the unserved energy requirements if sub-problems were infeasible. Those cuts, from mathematical perspective, were Lagrangian duals indeed; but from engineering angle, they meant the costs associated with 1MW increment or decrement. However, the network constraints were only dealt with by maximum or minimum transmission capacity, although this way greatly simplified the decomposition effects.

2.2.6. Fuzzy Logic or Knowledge-based model

Fuzzy logic, as its name shows, is not based on precise mathematical descriptions as traditional manners. It mimics the ways of people’s thinking. Instead of quantifying one uniform “value” to every uncertain variable, it uses the means similar to the linguistic declarations to give a threshold for every uncertainty extents. Its evaluation is associated with membership functions and fuzzy sets. Therefore, this manner is very natural to integrate expert rules into the model. Fuzzy logic and knowledge-based models were widely applied to traditional maintenance problems, which originated from the facts that sometimes the precise evaluations of some uncertainties were impossible. For example, the reliability indices would sometimes be barely determined accurately because of lack of valid and trustable data.
Moreover, fuzzy logic itself doesn’t contain any optimization sense and thus, if one needs that the results with optimality meaning; certain optimization technique must be integrated appropriately. Otherwise fuzzy models only give the “crisp values” by expert rules in terms of special concerns.

The general disadvantages of this method stem from the assumption that the behaviour of the stochastic variables must be presented or expressed by regular fuzzy set functions. This is not fully true because the performances of some uncertain variables are very unpredictable and hardly describable. Here are some representative papers:

Sergaki and Kalaitzakis (2002), standing on the level of plant components, gave an integral model to describe the system safety by the maintenance planning. This model used a fuzzy relational database for manipulating the data required for criticality component ranking in inspection planning of thermal power systems, incorporating criteria concerning various aspects such as safety, reliability, economy, operational conditions and environmental impacts. The most contribution was their proposed mechanisms of integrating complex knowledge bases, i.e., classification, composition and generalization respectively. Based on the specific and elaborate system definitions, fuzzy sets, membership functions associated expert rules were naturally established to reflect the impacts by signal component inspection/maintenance. This model was not in any sense of optimization.

Huang (1997) has the similar model structure with Sergaki’s. Though they both used fuzzy membership functions to present random natures of parameters, one different point is, instead of fuzzifying only two parameters (e.g. price and load), Huang tried to fuzzify all uncertain parameters including reserve, manpower, start timing and admissible maintenance. GA (Genetic Algorithm) was employed to enumerate all those random variables and attempted to make better offspring after crossovers and mutations. This is one special place because GA here was more like a simulation rather than an optimization strategy. The optimal decision was implemented by so called fuzzy dynamic programming, which would be mapped with the highest membership value, obtained from fuzzification process, for the intersection of fuzzy sets at the last stage. But one of obvious flaws of this paper was the suitability of assumptions on those membership functions. It is hard to believe that the random variables would perform such regular functional behaviours.

El-Sharkh and El-Keib (2003) used fuzzy sets to handle the uncertainties of load and fuel prices, associating evolutionary programming (EA) to achieve a near-optimal maintenance scheduling including the transmission maintenance constraints. For implementing the EA easily,
the model was separated into two parts, one was for maintenance scheduling and the other was for power dispatch problem. EA firstly generates lots of random enumerations, only those satisfying all maintenance constraints could be input into the power dispatch problem where load and fuel prices were treated as fuzzy membership functions. By certain criteria of comparing and selecting the individuals, EA could mutate better offspring and therefore converge to the optimal scheduling. The model finally gave a range of objective values (in the paper they were presented as maintenance costs) that described the random extents or ranges due to the uncertainties of loads and fuel prices. However, using fuzzy sets to do so seems not preferable, for the dynamic behaviours of those stochastic variables are too complicated to meaningful, if only described by a single membership function.

Mohanta et al (2005) solved a problem concerning the randomicity of FORs (forced outage rate). Traditionally, thermal plants reliabilities were measured by expect FORs that were based on the mean time to failure (MTTF) and the mean time to repair (MTTR). All of these parameter values were either from historical database or from some experts’ expectations, thus quite questionable from pragmatic perspectives. Trying to avoid using those expected (crisp) values and involving all other non-random uncertainties related to the mean time to failure/repairs, fuzzy sets became the only proper institution to evaluate LOLP based on those random FORs, which will revise the overconfidence effects by insufficient data collection as well as its’ vagueness and impression. After computing FMTTR and FMTTF, (i.e. fuzzy MTTR and fuzzy MTTF), transition probabilities in fuzzy sense could be explicitly defined on both success and failure states. Finally, FLOLP (fuzzy LOLP) would be determined by this fuzzy Markov model. The hybrid GA/SA technique was adopted for optimizing the maintenance timing. This paper was a very enhancement for better engineering approximation.

2.2.7. Heuristic / Genetic Based Algorithms

Heuristic algorithm has very wide applications in solving maintenance optimization model. There are several main domains:

1) Reliability Analysis. Lot of analysis based on reliability indices (i.e. LOLP) are trying to use heuristic methods. The general way is to find heuristically the maintenance duration for the unit with higher priority under the target of getting lower reliability indices. Then fix this duration and go on to the next unit. These ideas were often introduced in the papers regarding reliability concerns.
2) Special mathematical research. Some papers are of only theoretical purpose. Their authors tried various new algorithms but still in terms of conventional maintenance formulations, and intended to testify and prove the advantages of their novel ideas;

3) Multi-objective problems. Heuristic algorithms are also needful for multi-objective problems, which require kind of balances between various objective functions.

4) Genetic-similar algorithms. Genetic algorithms, essentially belonging to heuristic searching strategy are valid and efficient optimization techniques. There are quite lots of papers/applications based on variants of those optimization algorithms. Therefore those algorithms should also be grouped into this heuristic category.

One of the general deficiencies in this method category is their oversimplified constraints or considerations. From my personal view, many published applications were only from intentions of testing algorithm efficiency and most of them were based on quite simple formulations. As a result, although the outputs were sometimes revealed so successfully, they were not fully suitable to real cases. Another disadvantage is that they are short of valid foundations supporting the proof of results optimality. As we know that heuristic searching is always in terms of random enumerations. The performance of those random values can hardly be explained or controlled by conventional mathematical ways. For example, it is not easy to ensure the convergence of standard deviation for enumerated variables. Following are the brief introductions to some applications of the heuristic strategy.

Kralj and Petrovic (1995) reported their heuristic strategy for searching ‘relatively’ optimal value for multiobjective problems. MS problem has many different concerns, in terms of different standpoints of decision-makers. But it is usually not easy to integrate and balance them together. The authors proposed one algorithm, associated with the optimality testing procedure that can balance the trade-off among multiple objectives, like fuel costs, expect unserved energy and constraints violations in this paper. But this algorithm was dependent on simulations that need enumerations over all possible units’ outage combinations. As mentioned by the authors, the task difficulty would increase exponentially when the unit numbers increase. Thus its applicability is restricted in real surroundings.

Frost and Dechter (1999) mainly presented their mathematical interests in their paper. Their idea originated from an approach called backtracking beads constraints techniques. Backtracking is following such an idea as subsequent steps. Firstly consider and pick up one variable at a time from the problem, then evaluate this variable with the value from its domain that doesn’t violate any constraints, if there is no non-conflicting value available, then a dead-end
occurs. Under this circumstance, the algorithm tracks back to the previously instantiated variable and would select a new value for it. Otherwise treat this value as a valid assignment for allowing the optimal decision in the future. The backtracking algorithm, essentially as a special and efficient enumerating procedure under the constraints, traverses the feasible space of partial assignments in a depth-first manner. This kind of backtracking algorithm may gain some efficiency in searching an optimal value, but it needs delicate encoding skills and more importantly, it must convert particular variables from original formulas into the variable sets special for algorithm needs. Even so, the algorithm itself is still attractive if it can be associated with other dynamic schemes.

Digalakis and Margaritis’s paper integrated both GA and parallel computation. Their approach’s name was also called parallel co-operating cultural algorithm (PARCA). In the network of workstations, a master processor is in charge of creating the initial population, managing the population, performing selection, recombination and mutation. When the solutions need to be evaluated, they are dispatched to the salve processors which manage their own executions. Once each slave processor has carried out its execution experiments, the results are returned to the master processor. More efficiently, sub-populations are created to facilitate the management of population in each slave processor. Parallel strategy is definitely very promising in large scale, multi-components computations, but its implementation is also relatively difficult.

The paper first based on the simulated annealing methods proposed by Satoh and Nara (1991). This method was essentially in terms of enumerating random numbers as many as possible and in virtue of certain mechanism, the searching procedure can always approach toward convergence. As one common shortcoming of those global searching algorithms, random-enumeration-based methods are generally not straightforward for constrained problems. Thus penalty function is a usual manner to prevent the constraints being violated so much. “Annealing” is named after the thermal process of obtaining low energy states of a solid in a heat bath. In a general sense, it means that a steady process will forever go from high energy status which represents basically the disorder and inefficiency, into the low energy stage with very regular and ordered behaviours. “Temperature” is the indicator of convergence process. Simulated annealing could be understood as the process that system goes from chaos into stability. The advantages of this kind of methods are as obvious as their flaws. They are readily to be modelled and coded, but also quite vague in physical sense and no strictly proofs for optimality convergence.
Dahal and McDonald (1997) reported their GA methods in solving MS problem. They chose their objective function as the minimization of the sum of squares of the reserve, which was unusual in general. The focus of the paper concentrated on the encoding ways of GA algorithm, namely by binary, binary for integer and integer forms. Those three encoding forms are testified, and compared in terms of their performances. Conclusions showed that integer encoding had many advantages, for example, it was more intuitive, efficient and robust. Furthermore, it seemed easier for implementation.

Kim, Hayashi and Nara (1997) brought us a very useful and interesting paper based on the integration of genetic algorithm (GA), simulated annealing (SA) and tabu search (TS). The idea of combining SA and traditional GA comes from the observations that the convergence performance of GA can be improved by taking the probability of SA as one criterion when accepting new trial solutions. Although the algorithm efficiency and accuracy seemed to be improved, its possibility of dropping into the local optimum was also becoming larger. Under this circumstance, tabu search was consequently brought into attention for handling this problem. Tabu search can search through the neighbourhood regions of the solution and take them as genetic string. Moreover, the intensification and diversification of the searching regions representing the long-term and intermediate-term memory of TS have also been proved to be very efficient in strengthening the searching ability. The final results show that this integrated searching mechanism by GA, SA and TA do perform better in most instances, and seem very necessary to replace the traditional combinatorial optimization techniques.

Burke and Smith argued in 2000 about efficiency results of simulated annealing reported by Satoh and Nara (1991) in comparison with other genetic or evolutionary methods. The structure of the paper was quite similar to that of the one by Kim et al. (2000). Aiming at the efficiencies of different evolutionary approaches and local search algorithms, they were trying to give a full comparisons and descriptions. The maintenance model itself was following Satoh and Nara’s except the so-called “wrap-round maintenance periods” included. A very attractive contribution in the paper was that the authors combined different techniques together and formed hybrid algorithms that were tested and contrasted against every single part/component alone so as to investigate its effectiveness. For example, the paper showed that simulated annealing plus tabu search is better than annealing alone but worse than tabu search. In addition, an approach called memetic algorithm that modelled culture evolution, was successfully designed and applied in comparison with the results by those conventional evolutionary approaches. The final
conclusion revealed that a memetic algorithm using tabu search as a local optimizer produced the best results and was much better than simulated annealing algorithm reported by Satoh and Nara.

### 2.2.8. Miscellaneous Coordination Techniques

It is obviously a trend that some researchers started thinking the maintenance strategies in a deregulated or competitive environment. Wherein, independent system operator (ISO) must coordinate the unit’s outage strategies among all generation companies (GENCO) in order to satisfy certain reliability or safety requirements. Inversely, every GENCO concerns more about the profits coming from competitions. As a result, the coordination techniques were coming up necessarily. The main start point of those coordination techniques is still from the balance concerns. But differently the new balance is based on the relations between ISO and GENCO.

Conejo (2005) focused on the relations between ISO and each producer in restructured electricity marketplace and hence proposed a mixed-integer model to negotiate the bilateral concerns, namely reliability and profit. In other word, this model tried to find a trade-off between the highest security and best benefit. In details, the model had two parts; one was called MR-MS (maximum-reliability maintenance scheduling) and the other was MP-MS (maximum-profit maintenance scheduling). ISO would run MP-MS firstly and proposed a plan usually associated with highest grid security. Then each producer would run MP-MS individually. If the results of the two parts were close enough, ISO would adopt them as an optimal schedule. Otherwise an institution called incentives/disincentives would be applied by increasing the profitability at the time windows when the price was not attractive enough, so as to “persuade” producers to change their schedules. Essentially this idea was similar to a multi-objective problem. As a kind of heuristic descent method, several iterations were necessary to make MP-MS and MR-MS convergence. The disadvantage of this model lies in the fact that it doesn’t contain any uncertainties so far. Also, too many binary variables may extend severely the running time.

Billinton and Mo gave at the same year (2005) another method called MCT, abbreviated for Maintenance Coordination Technique. Differently, the authors draw supports from commercial software named MECORE, which was reportedly enhanced by BC Hydro. MECORE was Monte Carlo-centered composite system reliability evaluation tool. The basic procedure of MCT was that, at the beginning, ISO would receive all the outage plans from entire GENCO and TRANSCO (Transmission Companies). Then ISO would run MECORE whose purpose was to enumerate or simulate all system components states and calculate annualized reliability indices (EENS, for instance) at system peak load level. If the calculated reliability
indices were below the acceptable level, then the outage was approved, otherwise ISO would recommend GENCO to delay outages until system safety would be guaranteed. The paper also discussed the importance of one method when system risks were acceptable but load point risks were not. Generally, this kind of outage planning was not in the real sense of optimization, just from pure risk analysis. In addition, Monte Carlo simulation also seemed to be sort of compromise because the composite system consisted too many unknown states. Although so, this paper still pictured the future developments of maintenance planning strategy in deregulated industries.

2.2.9. Game Theory

Game-theory-based papers have not become quite popular until recent years. Using game theory Kim (2003) published the first paper in which he tried to explain the optimal strategies of maintenance scheduling in a competitive marketplace by different GENCOs (generation companies). However, game theory is not well developed, not only due to the incomplete theory itself, but also due to deficient understandings of market behaviours. Fundamentally, game theories are established on a concept, also called Nash Equilibrium, which indicates the equilibrium condition showing that one player can get the best profit if he can’t gain any additional benefit increments by changing his own strategy while all other players keep their own strategies unchanged. Since solving Nash Equilibrium is not an easy thing, Kim used an approach based on exhaustive searching, which helped him compute the profits for all strategies satisfying Nash Equilibrium. Obviously this way is only possible for very small scale problem. Also their model had a very strong assumption that stated that each player would know explicitly any information from other players or rivals. This assumption is not suitable for the competitors in current deregulated marketplaces, because for most of time, people are making strategies under imperfect outside knowledge. Moreover, the reliability constraints are not considered in the game-theory-based model.

Chattopadhyay (2004) continued the game theory’s application, but his research was concentrated on a very special case, i.e. oligopolistic marketplace, where a so-called Cournot-Nash Equilibrium could be used for modelling the strategy relations among all the GENCOs and their corresponding intertemporal effects. This Cournot-Nash Equilibrium is valid only under the situation that the GENCOs have the ability to influence market prices by altering their generation and maintenance decisions. They would like to exercise such market power to attain their profitability; therefore, from this fact, the decisions for all GENCOs are interrelated. The model
was mainly developed from Cournot generation strategy, including the considerations of perfect competition, firm costs and effects of oligopolistic competition. The temporal effect was measured by the marginal value of capacity in given period and the marginal value of maintenance for each unit. The result showed that GENCOs would like to strategize their maintenance and generations decisions to maximize their profit and thereby keep prices above the short-term marginal cost. It is unlike the conventional model that mainly considered the economic opportunity for outages.

Although game theory in maintenance scheduling is quite undeveloped, it does reveal some properties that traditional models do not have. Without doubt, this direction will be focused further in the future when more and more generation companies are put into the competitive environments.

2.3. Maintenance Optimization Models in Other Industries

As discussed in the general overview section, quite a lot of industrial fields are putting more and more money and energy into the research of maintenance modelling. But in view of different operating characters of various industries, their maintenance strategies definitely have particular emphases and concerns. Here several representative papers are selected to show the different aspects of maintenance problem in contrast to electricity industry. According to the known published articles, many of them focused on component reliability analysis, associated with the costs caused by maintenance in the manner of either loss of production opportunities or their maintenance costing. Here is one example.

Vassiliadis and Pistikopoulos (2001) presented their work in chemical engineering by integrating availability, maintainability and reliability into an optimization model. Different from electricity utilities, the chemical system may have intermediate product if one or more units fail or are being maintained. That means the reliability or availability of each working unit should influence the total system profitability and consequentially, those traits must be considered appropriately in the systematic model. Also, due to the existence of multiple working situations induced by unexpected units’ failure, the system has therefore a lot of status, which will impact the final process revenue. The objective is to find a balance between the total revenue and maintenance costs. As consequence, all of these possible statuses have to be included into the objective functions. The second obviously different point is that the units in chemical engineering may be maintained several times during the study period (one year for instance). This differs significantly from the electrical utilities that don’t want to maintain frequently the
working equipments and commonly allow them to be out of service only once a year. Generally speaking, this model is too analytical and complex to be practical in reality because it belongs to a non-linear integer problem. Many simplifications have to be made so as to let the model solvable. As mentioned by the authors, the proposed solving strategy is not based on a rigorous mathematical background. Nevertheless, there are still many contributions valuable for electrical engineering, such as the idea of integrating the random maintenance together. In addition, in view of those potentially various statuses of working components, Markov decision model may probably become more preferable and suitable for numerical solutions.

Tan et al. (1997) developed a framework methodology also applied in the chemical engineering, but the reliability model, data collection and risk analysis problem were not concentrated there. Instead, the authors put forward a scheme proving that Monte Carlo simulation was a good way to deal with the uncertainties. By Monte Carlo simulation, large potential failure events can be “created” according to pre-assumed probability distributions. Then maintenance and lost production costs could be cumulated in order to give an overall estimate of total cost rate. According to central limit theorem, the expect cost can be taken equally as the arithmetic average value if the simulation goes to infinity. Another advantage of using Monte Carlo method is, it not only can handle easily various random variables, but also will increase the computational complexity linearly with more components being involved. In terms of those facts, a new modified genetic algorithm was proposed that would search the optimal “cost rate” by employing one Monte Carlo simulation. However, in a general context Monte Carlo techniques are not believed to be trackable, but the authors believed that their approach could “tail” the simulation “tracks” after appropriate modifications. At the final part of this paper, the authors suggested that their proposed Monte Carlo scheme with genetic algorithm were general enough for common maintenance optimization problem.

At a strategic level, Winden and Dekker (1998) developed a Markov model for rationalizing building maintenance. This paper was enlightened from situation that government wanted to know the relations between the building qualities and minimum required level associated with maintenance costs. The building quality was usually measured by aggregated components, each of which has monotonous deterioration tendency with respect to time. The difficulty of this problem existed in one place, that is, renewing or repairing one component didn’t need to spend a whole time stage completely (such as one year). In other words, although there was a transition from one potential state to another, the connecting arc, if it represents the cost-to-go function in essence, didn’t necessarily mean the total maintenance cost because
another expenses, like residence costs had also been included during a typical time interval/stage. Those factors could induce the house maintenance problem into the category of semi-Markov decision process that should be transferred to standard Markov approach such that it can to be solvable. The authors showed their simple approach by adding residence situation to the state space. But their means was too case-specific to be generally adopted by others.

Unlike chemical engineering, where the components reliability are highly concentrated on the impacts of system safety when scheduling the maintenance, the considerations in maintaining road facilities are much broader. For example, roads maintenance would consider different pavement segments, their deterioration functions, warning levels by distress types, maintenance or repair methods, costs, resources allocations and traffic loading, etc. Another specific character in road maintenance is that there are the options and alternatives between pavement rehabilitation and maintenance activity. If all those concerns are included into one model, the combinatorial complexity is too astounding to be tackled by current computers. Based on this challenge, Fwa, Chan and Tan (1996) put forward a GA based method, which showed the potentiality to solve this problem largely. They specified an integer encoding scheme that made the problem parameters very efficiently and successfully represented. Meanwhile the trade-off between rehabilitation and maintenance could also be balanced in an appropriate manner.
CHAPTER 3 MODELING METHODOLOGY AND SOLUTION ALGORITHM

3.1. Introduction

A maintenance scheduling model is essentially a combinatorial problem, which usually could be reduced into a time sequencing procedure. That is, when a system operator is determining units’ outages, his/her concern will be focused on determining the best appropriate time to perform the function of maintenance. For sure, this best time sequencing criterion, or so-called “optimality” could have several meanings. For example, in the BC Hydro system, it could represent the trade-off between importing/exporting or storing water in the system reservoirs. Taking units out for maintenance could potentially allow for the increase in imports opportunities. On the other hand, the import potential could also be reduced, as the system rotational energy will decrease and voltage stability issues over ride system efficiency objectives. These operational considerations poses an important question: what is the best method that could be used to integrate all of these considerations or constraints and determine the best time sequencing of maintenance for a particular generating unit in a large scale hydroelectric system?

It is well known that an algorithm based on mixed-integer programming could constitute a natural fit for this problem, where several integer or binary variables could be used to represent units’ status, more specially, using a 0 to designate a unit off status and vice versa, a 1 to represent a unit on. Also, other integer variables could be used to record the durations when the unit is on. All other operational requirements could be compiled into a set of constraints and the model could be formulated to be very intuitive.

As indicated earlier, GOM is a deterministic optimization model that is based on a linear programming algorithm and is extensively used by the B.C Hydro system operations engineers to help them in making planning and operational decisions. The GOM model was adapted and enhanced to solve the maintenance scheduling problem. The enhanced model was named the Generalized Optimization Model Maintenance Scheduling (GOMMS). But there is one inherent problem that needs to be addressed when GOM is adapted to solve the maintenance scheduling problem. Basically, modeling the production function in GOM is based on modeling the relationship at the plant level, where the units are assumed to be committed and loaded in an optimal fashion. The maintenance scheduling problem, however, should model the performance of the plant while considering how individual units are committed. This problem must be dealt with in order for a modified version of GOM to be used to solve the MS problem, and the methods and techniques for handling this issue will to some extent decide the efficiency of
GOMMS in solving practical problems. Finally, currently there are no uncertainties included in the model outlined below, and the GOMMS system will therefore be a deterministic mixed-integer programming model.

3.2. Mixed-Integer Programming Model

Listed below are the notation used and the formulation of the GOMMS system:

**Superscripts used**

- \( GN \) Total hydro plants,
- \( H \) The number of time step divisions,
- \( T \) Total time steps,
- \( M \) Total possible combos in one plant,
- \( N \) Total segments of piecewise linear functions.

**Subscripts used**

- \( t \) The \( t^{th} \) time step
- \( j \) The \( j^{th} \) hydro plant
- \( h \) The \( h^{th} \) sub-time step
- \( n \) The \( n^{th} \) segment of one piecewise-linear function
- \( m \) The \( m^{th} \) combo index
- \( u \) The \( u^{th} \) unit index
- \( nt \) The \( nt^{th} \) start timing (week) for certain unit
- \( b \) The \( b^{th} \) base measuring the combo value by equation (3-49)
- \( r \) The \( r^{th} \) operable region in certain HPG curve
- \( z \) The \( z^{th} \) possibly combined operable zone in certain HPG curve

**Parameters**

- \( ABMax_{j,t}/ABMin_{j,t} \) The allowable max/minimum exports/imports to Alberta at \( t^{th} \) time step
- \( BCombo_{j,m,b} \) Vector parameter, representing the \( m^{th} \) combo value at the base \( b \)
- \( D_{j,u} \) Encoding of unit \( u \) at plant \( j \)
- \( FBbkp_{j,m,n} \) Forebay increment point at the \( n^{th} \) HPG segment, \( m^{th} \) combo of plant \( j \)
- \( FBMax_{j,t}/FBMin_{j,t} \) The allowable max/minimum elevation in the \( j^{th} \) reservoir at \( t^{th} \) time step
- \( G_{ORO,j} \) Operating reserve obligation fraction (e.g. 0.05)
- \( Load_{h,t} \) The pre-specified load value at time \( t \), sub-time step \( h \)
- \( MD_{j,u} \) The outage duration of unit \( u \) in plant \( j \)
- \( ME_{j,u} \) The pre-specified date when the outage of unit \( u \) must end before that
- \( MS_{j,u} \) The pre-specified date when the outage of unit \( u \) must start after that
- \( NetLimit_{t} \) Transportation limit at time \( t \)
- \( ONES_{j,u,nt,t} \) The status of unit \( u \), evaluated by binaries if its outage starting since the \( nt^{th} \) week
- \( OutageTiming_{j,u,nt,t} \) The status of unit \( u \), evaluated by its type encoding if its outage starting since the \( nt^{th} \) week
\begin{align*}
P_{\text{Exports},t,h} & \quad \text{The pre-scheduled exports at the } t^{th} \text{ time step, the } h^{th} \text{ sub-time step} \\
P_{\text{Imports},t,h} & \quad \text{The pre-scheduled imports at the } t^{th} \text{ time step, the } h^{th} \text{ sub-time step} \\
P_{bkp,j,m,n,h} & \quad \text{Generation breakpoint at the } h^{th} \text{ forebay, } n^{th} \text{ HPG segment, } m^{th} \text{ combo of plant } j \\
P_{\text{Intercept},j,m,t} & \quad \text{The intercept of efficiency function under combo } m \text{ in plant } j \\
P_{\text{Lbkp},t} & \quad \text{The left output limit on the } r^{th} \text{ operable segment of certain HPG curve} \\
P_{\text{Maxbkp},j,m,n,h} & \quad \text{Maximum P breakpoint at the } h^{th} \text{ forebay, } n^{th} \text{ HPG segment, } m^{th} \text{ combo of plant } j \\
P_{\text{Max},j,m,t} & \quad \text{The available maximum output in plant } j \text{ at time step } t \text{ with unit combo } m \\
P_{\text{oth},h} & \quad \text{The energy generated by other resources at the } h^{th} \text{ sub-time step} \\
P_{\text{Rbkp},r} & \quad \text{The right output limit on the } r^{th} \text{ operable segment of certain HPG curve} \\
P_{\text{PRE},t} & \quad \text{Pre-scheduled rotational energy at time } t \\
Price_{\text{AB},t,h} & \quad \text{The spot price in Alberta spot market (CDN $)} \\
Price_{\text{US},t,h} & \quad \text{The spot price in US spot market (US $)} \\
Q_{bkp,j,m,n,h} & \quad \text{Discharge breakpoint at the } h^{th} \text{ forebay, } n^{th} \text{ HPG segment, } m^{th} \text{ combo of plant } j \text{ (see fig. 1)} \\
Q_{IR,j,t} & \quad \text{The natural inflow into the } j^{th} \text{ reservoir at the } t^{th} \text{ time step} \\
Q_{\text{Maxbkp},j,m,n,h} & \quad \text{Maximum Q breakpoint at the } h^{th} \text{ forebay, } n^{th} \text{ HPG segment, } m^{th} \text{ combo of plant } j \\
Q_{\text{PMax},j,t} / Q_{\text{PMin},j,t} & \quad \text{The allowable max/minimum discharge from the } j^{th} \text{ plant at } t^{th} \text{ time step} \\
Q_{\text{Flash},j,t} & \quad \text{If there is flash requirement of the } j^{th} \text{ reservoir, } t^{th} \text{ time step, otherwise } \infty. \\
Q_{\text{Max},j,t} / Q_{\text{Min},j,t} & \quad \text{The max/minimum discharges from turbines in the } j^{th} \text{ plant at } t^{th} \text{ time step} \\
r & \quad \text{Exchange rate of CDN $ to US $} \\
R_{\text{Combo},j,m,t} & \quad \text{Equivalent representative combo, evaluated by equation 3.31 at the } m^{th} \text{ combo of plant } j \\
R_{c,t} & \quad \text{The intercept of linear function in (fig.3-6) in terms of load at time step } t \\
R_{m,t} & \quad \text{The slopes of linear function in (fig.3-6) in terms of load at time step } t \\
R & \quad \text{Reserve ratio at time } t \\
S_{R,t,h} & \quad \text{The minimum spinning reserve requirement at } t^{th} \text{ time step, } h^{th} \text{ sub-time step} \\
U_{RE,j,u} & \quad \text{Unit rotational energy specified by BCTC} \\
U_{\text{Max},j,m,t} / U_{\text{Min},j,m,t} & \quad \text{The allowable max/minimum exports/imports to US market at } t^{th} \text{ time step} \\
Y_{C,j,m,u} & \quad 1 \text{ if unit } u \text{ is on in plant } j \text{ under combo } m, \text{ otherwise } 0. \\
Y_{z,r} & \quad \text{The 0-1 array specifying the current running segment of certain HPG curve} \\
\textbf{Variables} \\
F_{B,j,t} & \quad \text{The elevation of reservoir } j. \text{ Generally } F_{B,j,r} \text{- } V_{j,t} \text{ is expressed in a piecewise linear form.} \\
8_{m,t} & \quad \text{Binary variable specifying the combo in the } j^{th} \text{ plant at the } t^{th} \text{ time step.} \\
P_{j,t,h} & \quad \text{The energy generated by the } j^{th} \text{ plant at the } t^{th} \text{ time step, the } h^{th} \text{ subtime step} \\
Q_{S,j,t} & \quad \text{The spill discharge from the } j^{th} \text{ reservoir at the } t^{th} \text{ time step} \\
Q_{T,j,t} & \quad \text{Total units' discharge from the } j^{th} \text{ reservoir at the } t^{th} \text{ time step} \\
Q_{TM,j,m,t,h} & \quad \text{Total units' discharge with combo } m \text{ from the } j^{th} \text{ reservoir at the } t^{th} \text{ time step} \\
\text{Spot}_{\text{AB},t,h} & \quad \text{The energy sold into Alberta spot market at the } t^{th} \text{ time step, } h^{th} \text{ subtime step} \\
\text{Spot}_{\text{US},t,h} & \quad \text{The energy sold into US spot market at the } t^{th} \text{ time step, the } h^{th} \text{ subtime step} \\
U_{D,j,u,t} & \quad \text{Binary variable specifying the action of shutting unit } u \text{ down} \\
\end{align*}
$U_{j,u,t}$ Binary variable specifying unit status in the $j^{th}$ plant at the $t^{th}$ time step.

$US_{j,u,t}$ Binary variable specifying the action of starting unit $u$ on plant $j$, step $t$.

$V_{j,t}$ The storage of the $j^{th}$ reservoir at the $t^{th}$ time step.

$\theta_{j,m,n,f,b,t}$ Weighting factors on the grids of 2D function $P(FBbpk, Qbpk)$ at plant $j$ and time $t$.

$\lambda_{j,m,n,f,b,t}$ Weighting factors on the grids of 1D function $PMax(FBbpk)$ or $QMax(FBbpk)$.

$\omega_{j,m,n,f,b,t}$ Binary variable indicating the segment index on HPG curve (2D case).

$H_{j,f,b,t}$ Binary variable indicating the forebay index (1D case).

$e_z$ Binary variable specifying the choice of running at the $z^{th}$ operable zone.
Equation (3.1) represents the water balance for reservoir $j$, and it simply states that the water inflows and outflows into a reservoir should be balanced at any time interval. Here, the subscript $k$ means all upstream reservoirs are hydraulically linked to reservoir $j$ directly. Constraint 3.2 states that at any time, the reservoir elevation should be within a certain range. In both GOM and the GOMMS systems, the relationship between $FB$ and $V$ is represented by piecewise linear functions. Also, in some studies the $FB$ levels in the last time step are set to a specified level, thereby forcing the reservoir to end at a certain levels.

### 3.2.2. Load-Resource Balance, Spinning Reserve and Market Limit Constraints

\[
\sum_{j} P_{j,t,h} + Poth_{t,h} + \text{P_{Imports}}_{t,h} - \text{P_{Exports}}_{t,h} - \text{Spot}_{US,t,h} - \text{Spot}_{AB,t,h} \geq Load_{t,h} \tag{3.3}
\]

\[
\sum_{j} \sum_{m} (g_{j,m,t} \cdot PMax_{j,m,t}) - \sum_{j} P_{j,t,h} \geq SR_{t,h} \tag{3.4}
\]

\[
\text{USMin}_{t,h} \leq \text{Spot}_{US,t,h} \leq \text{USMax}_{t,h} \tag{3.5}
\]

\[
\text{ABMin}_{t,h} \leq \text{Spot}_{AB,t,h} \leq \text{ABMax}_{t,h} \tag{3.6}
\]

Constraint 3.3 is the load-resource balance equation and states that the total generation by target plants ($P_{j,t,h}$) and all other resources ($Poth_{t,h}$), which is assumed to be a known, plus imports minus exports, should be greater than or equal to domestic load at any time step. Constraint 3.4 follows Chang, Gary W et al (2001) formulation, and is called the spinning reserve constraint. It physically requires the system to hold certain amount of spinning reserve for system reliability concerns. The constraint hedges against the possibility that too many units are out of service simultaneously and provides the reserve needed to maintain system stability. Here index $m$ corresponds to a unit combo for plant $j$, and binary variable $g$ indicates that at any time, plant $j$ could only have one combo ($g=1$). It will be shown in subsequent sections that the $g$ actually provides a link between a plant and its units. This constraint could be optional. Section 3.2.6 will provide more specific modelling details on rotational energy requirements as set by BCTC. Constraints 3.5 and 3.6 limit the amount of energy trade in the US and the Alberta electricity markets. Negative values indicate that the system is exporting energy, and positive
values indicate the system is in import mode. In addition, a total domestic load value is used in both GOM and GOMMS, which represent the sum of load for all regions in BC.

### 3.2.3. Plant Generation and Turbine Flow Constraints

\[
\sum_{m} g_{j,m,t} \cdot QTMin_{j,m,t} \leq QT_{j,t,h} \leq \sum_{m} g_{j,m,t} \cdot QTMax_{j,m,t} 
\]  

(3.7)

\[
QPMin_{j,t} \leq (QT_{j,t,h} + QS_{j,t,h}) \leq \min \{QS_{Flash_{j,t}}, QPMax_{j,t} \} 
\]  

(3.8)

\[
PMin_{j,t} \leq P_{j,t,h} \leq PMax_{j,t} 
\]  

(3.9)

\[
\sum_{m} [P_{m}(QM_{j,m,t,h}) - P_{intercept_{j,m,t}}] \cdot (1 - g_{j,m,t}) \geq P_{j,t,h} 
\]  

(3.10)

\[
0 \leq QTM_{j,m,t,h} \leq g_{j,m,t} \cdot QTMax_{j,m,t} 
\]  

(3.11)

\[
\sum_{m} QTM_{j,m,t,h} = QT_{j,t,h} 
\]  

(3.12)

![Figure 3-1: Illustrative Hydro Plant Generation Curve under certain combo and forebay](image)

Figure 3-1 presents the piecewise linear function of the generation production function \( P \). Subscript \( m \) indicates that this function is indexed over the set of combos in a plant. In another words, each combo may have different production function \( P \). A set of binary variables \( g \) is designed to exclusively restrict the selection of the \( P \) function for a single combo in each time step. Equation 3.10 is a tricky as the combo index \( m \) doesn’t appear in the right hand side of the
equation, so summation over m is needed for the equation to be valid. Equation 3.11 restricts the turbine discharge to a range of values between zero and than maximum turbine limit for the selected combo in the optimization process. If a combo is not chosen, then the corresponding $Q_{TM_{j,m,t}}$ is set to zero. The corresponding plant generation however is not necessarily zero. To overcome this problem the plant generation curve is usually set to zero generation at zero turbine flow for such problems. Equation 3.7 requires the turbine discharge, $Q_T$, to be bounded by the maximum and minimum turbine limits. Since power generation is calculated as a function of the optimized turbine discharge, it does not need to be bounded, as the turbine bounds will be sufficient to satisfy these limits. Numerical tests show that the model listed in equations 3.7 - 3.12, can be solved in reasonable time.

![Figure 3-2: MCA HPG surface at combo 15](image)

Equation 3.10 indicates that the piecewise linear relation $P$ is only a function of discharge $Q$ for a given combo. This however is not true because $P$ should also depend on the forebay for a given plant, as illustrated in Figure 3-2. In the GOM model, the forebay level at each time step is assumed to be known, and it is given as a parameter in the model. Currently, the GOM model is run iteratively to update the forebay level in each iteration until the difference in forebay level converges to a small value. The GOM algorithm has a built-in mechanism to calculate an adjustment to the calculated turbine discharge if the assumed forebay level differs from the optimized water level. These adjustments accelerate the head convergence iterative process. These approximations give adequate results in practice as the forebay changes at every time step and between successive runs are usually small and their effect on the overall solution can be ignored.
The following formulas gives one possible modelling technique to handle the non-linearity of head dependence in the linear programming model outlined above. The key concern of dealing with non-linearity can be reduced to the issue of how to interpolate 1D and 2D surface functions by linear equations. Babayev (1997) was the first to present a potentially useful approach using mixed-integer programming algorithms to interpolate any arbitrary continuous function with one or two variables. Nevertheless, directly using the formulas he developed seems impracticable in the maintenance scheduling (MS) problem at hand. Therefore and taking the main ideas he has developed, and with making special considerations for the unique nature of the MS problem, equations 3.13-3.30 could potentially used to address the issue of variable head in hydro plants:

\[
P_{j,t,h} = \sum_{m} \sum_{n} \sum_{fb} P_{bkp \ j,m,n,fb \ \cdot \ \theta_{j,m,n,fb,t}}
\]  \hspace{1cm} (3.13)

\[
Q_{T, j,t,h} = \sum_{m} \sum_{n} \sum_{fb} Q_{bkp \ j,m,n,fb \ \cdot \ \theta_{j,m,n,fb,t}}
\]  \hspace{1cm} (3.14)

\[
F_{B, j,t,h} = \sum_{m} \sum_{n} \sum_{fb} F_{Bkpn \ j,m,n,fb \ \cdot \ \lambda_{j,m,n,fb,t}}
\]  \hspace{1cm} (3.15)

\[
P_{Max_{j,m,t}} \leq \sum_{m} \sum_{n} \sum_{fb} P_{Max_{bkp \ j,m,n,fb \ \cdot \ \lambda_{j,m,n,fb,t}}}
\]  \hspace{1cm} (3.16)

\[
Q_{Max_{j,m,t}} \leq \sum_{m} \sum_{n} \sum_{fb} Q_{Max_{bkp \ j,m,n,fb \ \cdot \ \lambda_{j,m,n,fb,t}}}
\]  \hspace{1cm} (3.17)

\[
P_{j,t,h} \leq \sum_{m} P_{Max_{j,m,t}}
\]  \hspace{1cm} (3.18)

\[
Q_{T, j,t,h} \leq \sum_{m} Q_{Max_{j,m,t}}
\]  \hspace{1cm} (3.19)

\[
\sum_{n} \alpha_{j,n,t} = 1
\]  \hspace{1cm} (3.20)

\[
\sum_{fb} \mu_{j,fb,t} = 1
\]  \hspace{1cm} (3.21)

\[
\sum_{m} \lambda_{j,m,fb,t} \leq \mu_{j,fb-1,fb,t} + \mu_{j,fb,t}
\]  \hspace{1cm} (3.22)

\[
\sum_{m} \sum_{fb} \theta_{j,m,n,fb,t} \leq \alpha_{j,n\cdot t} + \omega_{j,n,t}
\]  \hspace{1cm} (3.23)

\[
\sum_{n} \theta_{j,m,n,fb,t} = \lambda_{j,m,fb,t}
\]  \hspace{1cm} (3.24)

\[
\sum_{fb} \lambda_{j,m,fb,t} = g_{j,m,t}
\]  \hspace{1cm} (3.25)

\[
\sum_{n} \sum_{fb} \theta_{j,m,n,fb,t} = g_{j,m,t}
\]  \hspace{1cm} (3.26)

\[
\theta_{j,m,n,fb,t} \geq 0
\]  \hspace{1cm} (3.27)
\[ \lambda_{j,m,fb,t} \geq 0 \quad (3.28) \]
\[ \mu_{j,fb,t} \in \{0, 1\} \quad (3.29) \]
\[ \omega_{j,n,t} \in \{0, 1\} \quad (3.30) \]

For a detailed derivation of these equations, the reader can refer to the Appendix, where two propositions are given to show that the above linear interpolation scheme has very satisfactory approximation property. In the equations, \( P_{Maxb kp_{j,m,fb}} \) and \( Q_{Maxb kp_{j,m,fb}} \) are used to model the maximum power output and the maximum turbine discharge in plant \( j \) under combo \( m \) and forebay \( fb \). Those two values were derived from the SPUC (Static Plant Unit Commitment) model and they are also used in the GOM model. Figure 3-3 illustrates a typical relationship between the forebay level and the maximum plant generation output for different combos for the GMS plant. \( Q_{Maxb kp} \) have a similar pattern as those depicted by the curves for the maximum plant generation curves.

![Figure 3-3: GMS max outputs under different combo & forebay](image)

Figure 3-4 shows the configuration of the \( \theta \) variable and illustrates the scheme of using it as a weighting factor to evaluate the 2D equation. Figure 3-4 (a) clearly illustrates that there is a unique set of \( \theta \) variable for a given combos, and there is only one level that \( \theta \) can have nonzero value if the corresponding combo is selected by the optimization algorithm. At most four of the \( \theta \) values could be nonzero for a particular combo (see the dotted rectangle zone in figure 3-4 (b)), and the sum of \( \theta \) values over \( m \) and the different \( fb \) levels should be equal or less that sum of \( \omega_{j,n-1,t} + \omega_{j,n,t} \) as indicated in equation 3.23.
In Figure 3-5, the value of $\lambda$ has similar meaning as that for $\theta$, but in this case it represents the weighting factors for a 1D equation, i.e., the relations between allowable output, or discharge, and the forebay level under certain combo and plant combination.

![Diagram of Combo and Forebay Levels](image)

**Figure 3-4: The $\theta$ grids of $P$ with respect to $FB$ and $Q$**

In addition, equation (3.24) ensures that $\theta$ and $\lambda$ correspond to a feasible combination, as the weighting factors dispatched on two adjoining forebay levels must be equal. The essence of both the 2D and the 1D interpolation schemes are in terms of the forebay level. Equations (3.25) and (3.26) guarantee that only the chosen combo can have nonzero weighting factors.

![Diagram of $\lambda$ Configuration](image)

**Figure 3-5: the configuration of $\lambda$ and its evaluation scheme**
The binary variables \( \omega \) and \( \mu \) are designed to specify the target grid location. The difference between them is that \( \omega \) is for 2D case while \( \mu \) is for 1D case. Equations (3.20) to (3.23) indicate that only one grid index could be identified along all of the breakpoints.

Finally, several brief conclusions can be drawn before using constraints set 3.13 to 3.30. The advantage of formulating the problem in such a manner is obvious. That is, the model doesn't need to iterate for water head convergence. But the disadvantage is that the numerical tests on practical problems revealed that the solution efficiency is low. This is understandable as the number of binary variables involved becomes excessive. Furthermore, it is also necessary to realize that with or without the water head convergence requirement, the derived optimal maintenance schedule will be the same or nearly the same with the modeling methodology outlined above. Therefore if the modeling objective is to get the best outage timing schedule, then adding those constraints could be less meaningful as the computational needed to solve the problem would probably be too excessive for practical use. Therefore a special algorithm was designed that strikes a balance between a reliable outage schedules and water head convergence. The followings steps outline the suggested methods.

Step 1: Use the prescheduled water level at each time step, run GOMMS without constraints (3.13) to (3.30) and obtain the optimal outage timing;
Step 2: Fix this outage schedule, add constraints 3.13 - 3.30 and run the model again (this time water heads are variables). Calculate the “real” forebay trajectory for the fixed maintenance schedules.
Step 3: Update the forebay trajectory and search for new outage timing, if they are the same or approximately the same, then stop else go to step 4.
Step 4: Check the current iteration counter, if less than the specified total iterations allowed, then go to step 2, otherwise stop with a failure message.

3.2.4. Maintenance Constraints

First, appendix B contains the original maintenance constraints are introduced and several relevant concepts are discussed. These maintenance constraints were mainly adopted from Chang (2001, 2002), but they also have been proved to be quite inefficient for large scale studies with many time steps. In addition, it could be realized that too many redundant information are used in these formulas. For example, the duration of maintenance could be given or could be fixed in
advance, so once people figure out the maintenance starting date, the ending date and unit’s status at each time step the maintenance schedule could easily be determined.

In view of that, a new concept called representative (or similar) combos will be introduced to represents a class of combos which are essentially the same in terms of efficiency, capacity, etc. For example, in the GMS plant there are 10 units, with 3 different unit types: unit 1 to 5 belongs to type 1, unit 6 to 8 are type 2 and units 9 and 10 type 3. If the maintenance schedule is only allowed to take a maximum of two units out for maintenance at a given time, then there are a total of 56 unit combination possibilities. But among these combinations, there only 10 which are considered representative unit combinations. With the idea of representative combos, it can be seen that picking up representative combos will yield much more efficient solution efficiency than by choosing from ordinary combos. The following example illustrates the concept.

<table>
<thead>
<tr>
<th>Combo</th>
<th>Description</th>
<th>Combo</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1007</td>
<td>Unit 5 is off</td>
<td>1015</td>
<td>Unit 4 is off</td>
</tr>
<tr>
<td>1019</td>
<td>Unit 3 is off</td>
<td>1021</td>
<td>Unit 2 is off</td>
</tr>
<tr>
<td><strong>1022</strong></td>
<td>Unit 1 is off</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It can be clearly seen that any one of these five combos could be a representative combo because each of them represents essentially the same unit combination composed of four units from type 1, three units from type 2 and two units from type 3. If we don’t consider the SVC mode (units 1 to 3 in GMS are SC capable), all five combos will have exactly the same efficiency and capacity characteristics. Then the following question becomes obvious: what if we let the model only select from representative combos instead of all possible combos? If it is practicable, then one can foresee that the solution efficiency will be highly enhanced because the number of representative combos is much less than total combos (1023 in the case of GMS). The following outline one potential method that can be used:

- Don’t make a decision on the unit status at each time step. There are three implicative conditions that can help us to simplify the situation.
  1. The maintenance activity is usually set by weeks.
  2. The study duration (usually in multiple of years) is fixed. Or even more preferably, the maintenance season is known in advance.
  3. Each unit has a prescheduled length for outage.

Now it is quite simple to enumerate all possible maintenance timing. For example, if one unit will have a 3-week outage during one year span (52 weeks), then it has at most 50
possible timings to start the maintenance, i.e. from week 1 to week 50. If we start at the 20th week, then the unit status will be evaluated as 0 (off) from week 20 to 22 and is set to 1 in other weeks. Thereby, the start timings for all units' maintenance activities could be enumerated and determined before the optimization model is solved. The remaining job is to use this information as a parameter, called \textit{OutageTiming}, with the subscripts of plant \(j\), unit \(u\), index of possible start timing \(nt\) and the time step \(t\). Let \(\text{OutageTiming}_{j,u,nt,t}\) be zeros if unit \(u\) in plant \(j\) is having maintenance at time step \(t\) which starts at the \(nt^{th}\) starting timing.

After doing so, the original binary variables \(US_{j,u,t} \), \(UD_{j,u,t}\) and \(U_{j,u,t}\) should be cancelled or fixed. Instead, one new binary variable, \(XTiming_{j,u,nt}\), is included in the optimization model. It can take the value of one if the \(nt^{th}\) alternative timing is chosen for unit \(u\) in plant \(j\). Once the starting date is known, then this unit status in every time step will be easily determined. The main objective of this step is to compress the redundant information so as to cut down the number of binary variables.

- Choose representative combo instead of all possible combos. As discussed above, the numbers of combos are essentially the same. So picking only representative combos will make more sense and improve the solution efficiency. But the challenge now is how to find a new equation connecting the units' maintenance timings with representative combos.

In the previous section (3.2.4), the traditional combo formula was defined. Nevertheless, if one realizes that it is just one way to evaluate the units' combinatorial enumeration, then it is possible to define other new equations by which one can easily establish the connections between maintenance start timings and unit combinations, and the following outline a suggested method. Encode units with the same type by a uniform number. For example, GMS unit 1 to 5 are encoded by 1; unit 6 to 8 are encoded by 10 and 100 encode the remaining units 9 and 10. Then the combo definition can be formulated as:

\[
\sum_{u} U_{j,u,t} \cdot D_{j,u} \quad (3.31)
\]

\(D_{j,u}\) is the encoding for each unit, e.g. 1, 10 and 100 in the above example. The value calculated by this formula distinguishes only the combos among different representative combos. Using the example in table 3-1 and equation 3.31, all five combos, i.e. 1007, 1015, 1019, 1021 and 1022, will have one combo value 234. We can now give a new name, \textit{equivalent representative combo}, for this value, which differentiate it the from the original "representative combo"
On the other hand, the original formula to calculate the decimal equivalent to a binary number \( \sum U_{j,u,t} 2^{(u-1)} \) implies that number 1 represents a unit on status and a 0 represents a unit off status. Although this representation is natural and could easily be understood, it is still not practical for use in binary problems as the problem size becomes large. Indeed, if a unit is on, its status can be evaluated by any non-zero values. For example, we can use unit type encoding \( D_{jiu} \) to present unit on-line status. Now assume if GMS unit 6 is on, then its status is 10. Also if GMS 9 is working, it is 100, and so on. Therefore the linear equation 3.31 can potentially and directly connect representative combo with each unit's status. The preceding discussion give rise to the parameter \( OutageTiming_{j,u,nt,t} \) that records all values of each unit's status at every time step. If this parameter adopts the \( D_{jiu} \) encoding to represent units' on-line status, then a natural linkage between maintenance timing and representative combo can be perfectly built:

\[
\sum m g_{j,m,t} \cdot RCombo_{j,m} = \sum u \sum m XTiming_{j,u,nt} \cdot OutageTiming_{j,u,nt,t} \tag{3.32}
\]

\( RCombo_{j,m} \) is called the equivalent representative combo, calculated by equation 3.31. This can be combined with a set of constraints of binary exclusivity:

\[
\sum m g_{j,m,t} = 1 \tag{3.33}
\]

\[
\sum m XTiming_{j,u,m} = 1 \tag{3.34}
\]

Explicitly, (3.33) requires that only one combo could be chosen at each time step. The following paragraphs explain equation (3.32) in a detailed example for the MCA plant. The four MCA units consist of two unit types, two of each. That is, unit 1 and 2 are of type one, which has a \( D_{jiu} \) encoding evaluated at 10. In a similar way, unit 3 and 4 are of unit type two, and are given encoding type 100. In principle, \( D_{jiu} \) type encoding can take any value, but it is more preferable if the arithmetic product of unit numbers and the corresponding type encoding in previous type group will be less than the type encoding for the next group. For example, the GMS plant has three unit types, which could be encoded as 3, 17 and 34 respectively. Now, type one has 5 units and it satisfies the condition because 5*3<17. But, the second unit type group (which has three units) violates the condition as 3*17<34 is not satisfied. This method of encoding will be confusing because the model could never tell the case when 2 units from type two are off and the other case when only one unit of type three is off. It can be easily verified that all these two cases have the same \( RCombo_{j,m} \) value of 100. Therefore this requirement will guarantee that \( RCombo_{j,m} \) will be exclusively evaluated and it corresponds to its representative combo.
Back to the MCA example, suppose that the “optimal” units’ maintenance schedule is given in Table 3-4. Now the total study duration is 9 weeks. Number 0 still represents unit off-line status. Each unit will take 2 or 3 weeks in maintenance and it could start from any week. First, list all alternatives of possible maintenance schedule and evaluate the parameter \( \text{OutageTiming} \) for unit 1.

**Table 3-2: parameter OutageTiming for unit 1 of MCA**

<table>
<thead>
<tr>
<th>OutageTiming(_{MCA,1,nt} )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Intuitively, there are a total of 8 possible starting dates, so index \( nt \) could take a value in the range from 1 to 8. Then only one of the binary variables, \( X_{TimingMCA,1,nt} \), could be “chosen” and hence evaluated to one. Basically, this constraint (i.e. Equation 3.33) guarantees the exclusivity of the maintenance activity for the corresponding unit. Assume that the maintenance of unit 1 will start in the third week, then the parameter array shown in bold in table 3-2 will represent \( \text{OutageTiming} \) and the parameter \( X_{TimingMCA,4,1,3} \) will be set to 1. In addition, because unit 4 is of a different type and it usually takes 3 weeks to complete the maintenance outage and suppose its outage will not be scheduled until the fifth week, then table 3-3 lists the values for parameter OutageTiming. Due to the 3-week outage duration, unit 4 could only have 7 “slots” to start the maintenance activities.

**Table 3-3: parameter OutageTiming for unit 4 of MCA**

<table>
<thead>
<tr>
<th>OutageTiming(_{MCA,4,nt} )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Constraints (3.32-3.34) are much more efficient that original formulas (B.1-B.8, Appendix B), not only because the number of equations are significantly reduced, but also
from the facts that the binary variables are remarkably simplified. Finally, Table 3-4 gives the values for the selected unit combination using equation 3.32 and equation B.9. Based on the values calculated by equation 3.31, the relationship between the equivalent representative combo and units’ status could clearly be seen, which is much easier than using the original combo formula as listed in equation B.9.

<table>
<thead>
<tr>
<th>t</th>
<th>Unit 1</th>
<th>Unit 2</th>
<th>Unit 3</th>
<th>Unit 4</th>
<th>Original Combo value by (B.9)</th>
<th>Modified RCombo value by (3.31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

Assuming that a maximum of two units could be maintained at the same time, then using the old method would result in 11 combo possibilities. Using the new concept of equivalent representative combo, there are only 6 combo possibilities. The impact of reducing the number of combs will be more apparent with plants with large number of combs, such as the GMS. Suppose that the GMS plant can have a maximum of 3 units out at one time. Then using formula 3.32, there will only be 19 candidates of equivalent representative combo to choose from, instead of 176 combs if the old method is used.

Consequently, using a single constraint 3.32 the maintenance scheduling requirements are fully implemented and the solution efficiency is also greatly improved. A simple example can help explain how this simple modification has improved the computational efficiency of the maintenance scheduling model. Assuming that: i) A maximum of 3 units can be taking out at any instance at the GMS plant, while only 2 units can be taken out at PCN, MCA, and REV, and only one unit can be taken out at ARD; ii) The study duration is 52 weeks; iii) Each unit has a three weeks outage duration. Using (B.1-B.8), the total number of binary variables amounts to 14,816. Comparatively, using equation (3.32), the total number of binary variables is only 3,228, which is approximately 20% of the original number of variables. Numerical tests have also verified that this improvement in efficiency can be significant, particularly in large scale problems.

3.2.5. Reliability Requirements

As discussed in the chapter on literature review, the reliability requirement is a very complex concept to address, and it can be measured by several indices. For example, the
traditional reliability index LOLP is calculated by a probability convolution of the units’ forced outage rate (FOR) and loads. So if the maintenance scheduling for units has changed, the system LOLP would also change accordingly. To avoid this complexity, B.C Hydro uses a static reliability concept at the resource planning level, which could also be understood as system resource adequacy. This static reliability index is simply defined as a certain percentage (or ratio) of the system load at each time interval. In the BC Hydro system, the probability model for the LOLP determines that at the risk level of one failure day per 10 years, the system roughly requires a reserve of 14% of the total load in wintertime, or 10% in other time.

In addition, it is required that each plant should also have an operation reserve. This reserve can be explained as a conservative contingency in the case of a forced outage. In the GOM model, they are usually defined as a fixed ratio, e.g. 5%. Therefore, every plant must hold 5% of its full capacity for an operating reserve obligation to hedge against various risks. These two requirements can be easily modeled as a set of linear constraints as follows:

\[
\sum_j P_{j,t,h} + P_{\text{th}} + P_{\text{Imp}}_{t,h} - P_{\text{Exp}}_{t,h} - \text{Spot}_\text{US}_h - \text{Spot}_\text{AB}_h \geq \text{Load}_{t,h} \cdot (1 + R_t) \]  
\[
f_m(QT_{j,t,h}) \geq P_{j,t,h} \cdot (1 - G\_ORO_j) \]

3.2.6. Rotational Energy Requirement

The rotational energy requirement is primarily imposed for transmission system (voltage) stability purposes. Certain amount of rotational energy is used as a proxy for maintaining a minimum number of generating units on line to support the system voltage in periods of heavy import opportunities. The British Columbia Transmission Company (BCTC) computes and regulates the Must Run Requirement (RMR) for rotational energy at different system load levels. Figure 3.6.a illustrates the relationship between the system loads and the net transfer limits for different rotational energy levels (Mega-joules, or MJ).

Since the system loads are deterministic at each time step then the relationship between rotational energy and transfer limits could be calculated as shown in Figure 3.6(b). It can be seen that the relationship is linear. Using the variable, \( RE_t \), to represent the rotational energy at time step \( t \) and parameter \( URE_{j,u} \) to represent the rotational energy of unit \( u \) in plant \( j \), and parameter \( PRE_t \) to represent the pre-scheduled rotational energy at time step \( t \), and \( NetLimit_t \) to represent the total transfer limits at time step \( t \), then the system must satisfy limits:
Parameters $R_m$ and $R_c$ are the slopes and the intercepts of the linear relationship between the rotational energy and the net limits as illustrated in Figure 3.7. Once the relationship is plotted against load, the relationship of these equations can be easily observed as illustrated in Figure 3.7. It is then quite easy to derive the following regression equations as follows.

Using equations (3.39) and (3.40), the slopes and the intercepts can be determined based on the value of load at each time step in the study.
3.2.7. Synchronous Condenser

A synchronous condenser (S\C) is usually controlled by a voltage regulator to switch the unit to either generate or to absorb reactive power as needed by the system. The unit operates at full leading power factor and puts VARs (Volt-amperes reactive) onto the network as required to support a system's voltage or to maintain the system power factor at a specified level. Synchronous condenser units are also used for supporting voltage in situations such as starting large motors, or where power must travel long distances from where generated to where its used.

Simply speaking, S\C can work as a motor that absorbs or uses electricity from the grid to keep the runner rotating, hence providing an “economical” rotational energy that could support the system transmission capability. Because running S\C never consumes water from a reservoir or fuel from a thermal system, it can absorb reactive power from the transmission lines, and it is usually thought off as a cheap way to keep import/export capability of the system. This functionality is extraordinarily important during the maintenance season because running units on S\C could economically make up the losses in rotational energy as large units are taken out.

S\C units are not different from normal generating units when they run as generators. The only difference that can differentiate them is the generation production curve. S\C may add at least one negative output point on the generation axis. The following is an example from a S\C study for the MCA plant at BC Hydro, where unit 1 and 2 are currently S\C capable units. Suppose that unit 1 is off, and only unit 2 can work on S\C mode. It is known that S\C absorbs energy from the grid, so the generation production curve region can be represented by a point as shown in Figure 3.8 with negative generation (-8 MW) on the P axis. Region 2, 3 and 4 are
called operable ranges. Normally one turbine would not run along all discharge ranges because running in some of these ranges could cause some vibration problems. By combining all inoperable ranges for the available turbines, it is possible to know which discharge zones are operable ranges for a plant.

Through the use of binary variables the S/C condition could be formulated. Assume binary variable $e$ represents one of all the possible combinations of the operable zones shown in Figure 3.8 (as listed in Table 3.5). Parameters $PL_{bkp}$ and $PR_{bkp}$ represent the left and right output breakpoints of the operable segment $r$.

<table>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$ (m$^3$/s)</td>
<td>$P$ (MW)</td>
<td>$Q$ (m$^3$/s)</td>
<td>$P$ (MW)</td>
</tr>
<tr>
<td>$PL_{bkp}$</td>
<td>0</td>
<td>-8</td>
<td>94</td>
</tr>
<tr>
<td>$PR_{bkp}$</td>
<td>0</td>
<td>-8</td>
<td>106</td>
</tr>
</tbody>
</table>

It can be easily seen that the plant could be running in at least one of these operable segments in any time. As shown in Figure 3.8, there are several possibilities for running the plant in the operable zones:

1) Running at segment 1 and 2,
2) Running at segment 1 and 4,
3) Only in segment 3 (only for unit with S/C mode when it is running as generator),
4) In segment 4,

This information could be easily interpreted and stored into a parameter in the model as outlined below.

$$
Y_{4,*} = \begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}
$$

The row index $z$ in $Y$ specifies the state of combined operable zones. And the column index $r$ in $Y$ is used to indicate whether or not the $r^{th}$ operable segment is feasible. Therefore $\left( e_z \cdot Y_{z,4} \right)$ could be used to represent the current operable segments chosen by the model. The following formulas outline the formulation for this case:

$$
\sum e_z = 1 \tag{3.41}
$$

$$
\sum_r \left( \sum_z \left( e_z \cdot Y_{z,4} \right) \cdot PL_{bkp} \right) \leq P_{t,i} \leq \sum_r \left( \sum_z \left( e_z \cdot Y_{z,4} \right) \cdot PR_{bkp} \right) \tag{3.42}
$$
In this simple example only one S/C is available. If two or more S/Cs are available, it may create more complex combinations of operable zones, but form of equations 3.41 and 3.42 will hold. The two formulas are given for very specific combo. In such cases where the combo is a variable, then both $PLbkp$ and $PRbkp$ become variables. In this case the problem becomes much more intricate, although it can be handled by introducing more binary variables, but at the expense of making the problem much more difficult to solve.

It should be mentioned here that the S/C formulations outlined above will make the problem more difficult to solve, but the equations could easily be added to the model if the a quantitative impacts of S/C on maintenance schedules is needed.

3.2.8. Objective Function

Similar to the optimality criterion used in GOM, the GOMMS system also maximizes the value of exports from the BC Hydro system. The objective function in GOMMS is given in equation 3.43 below.

$$\sum_{t} \sum_{h} (Spot_{US,t,h} \cdot Price_{US,t,h}) + \sum_{t} \sum_{h} (Spot_{AB,t,h} \cdot Price_{AB,t,h}) \cdot r$$

(3.43)

The equations outlined in the previous sections constitute a basic mix-integer programming model to solve the maintenance scheduling problem of the BC Hydro system. Several equations can be dropped out to accommodate users’ specific needs, and this could results in many cases in improvement in model performance. The following sections will discuss further improvements on the integer programming part, but it is still preferable that the users of the model define the purpose of the study and see which constraints to include as they could have a major impact on performance of the solution efficiency of the problem.

3.2.9. Other Constraints or Considerations

In general the above formulation can implement the basic framework of a maintenance scheduling optimization model. Several additional or optional constraints could be added if the need arise. For example, GOM contains the monthly generation constraint that requires that the total monthly output from each plant be within a certain range. Other constraints could be added to account for transformers availability, such as in the case for the Peace Canyon (PCN) plant. In addition, GOM studies shows that to avoid a higher probability of spill in May and June, maintenance activities for the REV plant are usually maintained in times other than this period. It could be difficult to capture this requirement by GOMMS when the study is carried out in a weekly or even daily time steps. Therefore expert rules could be used to generate a set of constraints that can capture this and similar requirements, as will be further discussed in the
results chapter where it is shown that the maintenance schedule during this period will results significant reduction in the objective function than if the maintenance is carried out in the coldest month of the year.

Furthermore, in reality analysts prefer to arrange the maintenance activities in sequentially. Usually this kind of requirement results in a highly complex combinatorial problem and it is not easy to translate the requirement into a set of linear constraints. This problem becomes more complex for plants with many different units' types, particularly when the outage durations for these units are different. Certainly, the sequencing requirement problem can be partly solved by adding excessive number of binary variables, but so far it lacks a generalized formulation. Therefore in cases where there is no strict requirement to sequence the units maintenance for a given plant, the problem could be solved by another practical way. For example, numerical results shows that without the sequencing requirement, maintenance of units are usually clustered without adding specific constraints to force the sequencing, but in some cases with one or two weeks “break” in between. An easier and practical way would be to manually adjust the schedule for these units based on the experience of experts. The other way is to make the model run recursively by setting up several additional constraints for the next runs to try to avoid these “breaks” in the derived schedules. In the results chapter a method called “maintenance shape” was investigated as an alternatives approach if the “breaks” lasts more than 3 weeks or in cases where maintenance continuity is of high priority.

3.3. Simple Linear Relaxation Scheme

3.3.1. Brief Review of Linear Relaxation

The previous sections have shown that GOMMS is a very complex mixed binary/integer programming (MBIP) model with thousands binary/integer variables and constraints. Under some certain situations, such as in the case with finer resolution time steps, the solution process could become very lengthy. Meanwhile, sometimes very slight formulation changes could cause significant gains in the solution efficiency. The following sections outline some of the techniques that could be used to improve the performance of the solution algorithm.

As discussed previously, most of commercially available MIP codes adopt some form of a linear relaxation schemes to solve the MIP or IP problems. After relaxing the integer requirement, the original integer model becomes an easy to solve LP problem. Solving LP problems is much easier today using the highly developed LP algorithms, but the solution may contain some “fractional” values that seem meaningless. Therefore if there are some possible
ways which can force these fractional values to be integers then the solution process becomes very promising, as the model could be enlarged to contain either more specific constraints or finer time step resolution. The idea of using LP relaxation then naturally comes up, but in a different form from what is implemented in commercial codes. The LP relaxation scheme that was developed in this thesis is geared towards the engineering sense of the problem and takes advantage of the special structure of maintenance scheduling problem. Thus, no proof will be given to guarantee the global optimality of the solution, but numerical studies have indicated that it works efficiently and could attain acceptable results that could be implemented in practice.

The MIP problems that were solved by the linear relaxation algorithm were mostly based on simple and well-structured models. For example, Broughan and Zhu (2000) researched a well-structured integer model that appeared in the design of experiments. By analyzing the equality constraints, the authors were able to substitute them into the objective function. With several mathematical derivations, the authors were able to transform the original model to another type that can automatically guarantee the integrality of the solutions without including any integer variable in the model. The uniqueness of their case is from the simple and special structure of the objective function that is associated with integer variables, but this is not the case in the maintenance scheduling problem.

Trick’s (1991) presented a method that successfully employed the LP relaxation by a heuristic scheme to solve the generalized assignment problem. The generalized assignment problem is a special case of the transportation problem that belongs to $P$ hard problems. In these kinds of problems there are certain jobs to be assigned to certain amounts of machines, and each job can just be chosen by only one machine. On the other hand, there is a maximum total capacity for each machine. The objective is to get the maximum benefits or minimum costs through an optimal job allocation for the machines. It should be mentioned here that the MS formula 3.32 could also be explained in a similar manner to the assignment problem, because by appropriately allocating the unit maintenance activities, the operator would like to optimize the outage scenarios and achieve the best benefits. But the maintenance scheduling problem is much more difficult than the generalized assignment problem, because it not only involve the number and type of complex constraints, but the objective function is much more complex than the machine assignment problem as it involves complex hydroelectric and reservoir operations problem. It can be easily noted that no binary variables appear in the objective function, so the theoretical analysis presented by Trick’s cannot be adopted in the maintenance scheduling

63
problem. Nevertheless, solving the relaxed LP problem iteratively could still offer a good motivation to review his paper.

3.3.2. Using Cutting-Plane Method to Solve the LP Relaxation Scheme

Nemhauser (1988) in his book elaborated several relaxation schemes. Besides the linear relaxation scheme, there are also primal and dual heuristic algorithms, decomposition algorithms, including Lagrange (row) and Benders (column), and dynamic programming. Every algorithm has its own specific characteristics, so there couldn’t be an all-round approach covering all the cases. Appendix C briefly introduces the Fractional Cutting-Plane Algorithm (FCPA) from Nemhauser’s book. Simply speaking, FCPA is such an approach that is based on linear programming that adds the most frequently violated constraints into the model and effectively restricts the fractional values into integers. However, in the MS case this process is too complicated and it is not straightforward to use this approach as explained in Appendix C. But the idea of adding progressively valid cutting-planes is still helpful to “force” the relaxed variables to integer values.

The maintenance problem itself comprises of a rich mix of integer information that is hidden or implicit in the constraints. This gives rise to possibilities to reinforce the implicit integrality constraints and force the fractional values to become binary numbers. For example, the parameter \( \text{OutageTiming} \) is used either to specify the unit on-off status or the unit “capacity” which is represented by the unit type encoding \( D \). It is apparent that given different \( D \) values, the equality equation 3.32 will hold \( (R\text{Combo} \text{ will be matched by the corresponding } D \text{ values}). In integer programming, this information is redundant, but the LP relaxation algorithm it can serve as cuts to restrict the variables to integer values. Rewriting 3.32 in vector form:

\[
\sum m g_{j,m} \cdot R\text{Combo}_{j,m} = \sum u \sum nt X\text{Timing}_{j,u,n} \cdot \text{OutageTiming}_{j,u,n,f} \quad (3.44)
\]

The \( R\text{Combo} \) or the \( \text{OutageTiming} \) vector is evaluated in terms of different predetermined \( D \) values.

In BIP problems, the equivalent representative combo \( (R\text{Combo}) \) is designed to avoid excessively searching all the equivalent combos. But this screens out significantly useful integrality information. For example, knowing the current combo, it is easy to judge the status of each unit. But after using equivalent representative combo, this information is lost. For certain units, it’s the status of the units will not provide useful information. For this reason, equivalent representative combo is not suitable for the LP relaxation algorithm; instead a new parameter \(YC \)
(see equation B.2 in Appendix B) is adopted again, and correspondingly the constraint is modified as follows:

\[ \sum_m g_{j,m,t} \cdot YC_{j,m,u} = \sum_m XTiming_{j,u,m,t} \cdot ONES_{j,u,m,t} \]  

(3.45)

Parameter ONES is defined by 3.48, and it causes the unit "on" status to become an identity.

\[ ONES_{j,u,m,t} = \frac{OutageTiming_{j,u,m,t}}{D_{j,u}} \]  

(3.46)

\[ D \] is still the unit type encoding as defined in equation 3.31. Other implicit and useful integrality information stems from the definition of combo. Equation B.9 is used to calculate the value of combo by BC hydro now, and it assumes that the base is always 2. This equation can be extended to the case with any base situation. If using vector \( B \) to represent all the different bases and \( B_{Combo} \) for the corresponding combo, equation B.9 can be rewritten as,

\[ B_{Combo} = \sum_u U_{j,u} \cdot B^{(u-1)} \]  

(3.47)

which implies the integrality equality constraint as:

\[ \sum_m g_{j,m,t} \cdot B_{Combo \_j,m,b} = \sum_u U_{j,u,t} \cdot B_{j,b} \]  

(3.48)

Units' status \( U \) is determined by (3.34) or (B.2).

Although there are more implicit integrality constraints, like those in equations B.3 - B.6 in Appendix B, experience suggests that more cuts will also bring "fractional" convex hull too. So it becomes a balancing act between the cuts and the integrality requirements. Because the constraint cuts (3.44) belong to \( \pi \) (Cutting-Plane Constraint Sets), it should be added to the model runs. But the linear programming algorithm is already fast enough by the latest CPLEX version, so there is no difference between adding them one by one or all of them at the same time. Also, if implemented in a different way, then the coding work is much easier to be implemented. This method looks like a "brute force" method, but the facts show that this way works quite well in practice. Despite that the same integrality cuts are added, it should be noted that different formulations might have totally different effects on the results. That is, the key factors for forcing those "fractional" values to be integers could also be relied on in other constraints. For example, the generation formula

\[ \sum_m \left[ P_m \left( QTM_{j,m,t,h} \right) - P_{intercept \_j,m,t} \cdot \left( 1 - g_{j,m,t} \right) \right] \geq P_{j,t,h} \]

is not suitable to be included in the LP relaxation scheme, although it dose work very well when solved by MIP algorithms. If this formula is changed to the following format:
\( P_m(Q_{T,j,t,h}) + \text{lnf} \cdot (1 - g_{j,m,t}) \geq P_{j,t,h} \)  

then the fractional values will converge to integers very efficiently. But unfortunately, this formula converts the problem into a very hard to solve MIP problem and it will take exceedingly long time to solve with very slow convergence. Herein, a very large number named Inf is used to maximize \( P \) under the same \( Q_T \) in order to get rid of the combo subscript \( m \) at the right hand side of the formula. Specifically, the inequality sign will automatically maximize those reasonable power outputs with the same turbine discharge. In this case only the most efficient combo will be picked up in the optimized solution.

Although a large number of integrality cuts are added into the system, after several iterations the “fractional” values are still hard to converge to integers. So an important assumption has to be made. Because of the complexity of the MS model and as the iterative effects of the water head updates numerical experiments proved that the following assumption accelerates the convergence very effectively: the variable can be one if and only if its value in the current iteration is a fraction.

The assumption essentially says that if at each iteration the variables with a zeros value will be discarded in the next run. Intuitively, this seems to be a reasonable assumption because if a variable is already determined to be zero by the LP relaxation, then it has the least potential to be one as compared to others variables with factional values. In the MS problem, \( S \) represents the total set of variables and \( S_0 \) is the set of variables with zero values. So using the assumption, it has:

\[ \sum_{n \in S \setminus S_0} \text{XTiming}_{j,u,n} = 1 \]  

Finally, the scale of newly added cuts depends not only on the computer conditions, but also on the experience that was gained from running the current problem. It is not true that more cuts can definitely make the “fractional” values to converge to integer numbers faster. As discussed earlier, more cuts could bring a more complex convex hull, which could result in more “fractional” results. Experience suggests that for plants with multiple types of units, such as GMS, the vector \( \text{RCombo} \) in equation 3.44 can have members from 12 to 20, while the base vector \( B \) usually take values from 2 to 9. Although \( B \) could be a real number, from computational point of view it is better to have integer values because otherwise numerical errors could cause numerical instability.
CHAPTER 4 CASE STUDIES AND RESULTS

4.1. Case Study Introduction

To illustrate the efficiencies of different formulations and algorithms, the results from a one-year case study are analysed and presented in this chapter. The five largest plants in the BC Hydro system were included in this case study (Table 4-1). To understand the behaviour and performance of the algorithm in detail, the uncertainties of some of the parameters, such as inflow, market prices, and load, were not considered in this case study. Future research could address the impacts of these uncertainties on the maintenance scheduling problem. The following sections outline the inputs used in the study.

<table>
<thead>
<tr>
<th>Plant</th>
<th>Unit 1</th>
<th>Unit 2</th>
<th>Unit 3</th>
<th>Unit 4</th>
<th>Unit 5</th>
<th>Unit 6</th>
<th>Unit 7</th>
<th>Unit 8</th>
<th>Unit 9</th>
<th>Unit 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMS</td>
<td>261</td>
<td>261</td>
<td>261</td>
<td>261</td>
<td>275</td>
<td>275</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>PCN</td>
<td>175</td>
<td>175</td>
<td>175</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCA</td>
<td>435</td>
<td>435</td>
<td>465</td>
<td>465</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REV</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARD</td>
<td>95</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.1.1. Reservoir Nature Inflows

The water year selected for this study was 1964, which is often used as the representative year for GOM studies carried out at BC Hydro. The following plots (Figures 4-1 to 4-5) show the weekly average inflows to reservoirs that are included in this study.

![Figure 4-1: GMS weekly inflows in 1964](image-url)
Figure 4-2: PCN weekly inflows in 1964

Figure 4-3: MCA weekly inflows in 1964
Figure 4-4: REV weekly inflows in 1964

Figure 4-5: ARD weekly inflows in 1964
4.1.2. Reservoir Forebay Elevations and Limits

In the models that do not use elevations/heads as variables, forebay elevations and their monthly limits should be known beforehand. Figure 4-6 to Figure 4-8 present the pre-scheduled forebay trajectories and reservoir limits at the end of each month. Because PCN and REV have limited reservoir storage capacities, their operation depends heavily on the releases from upstream reservoirs, i.e., GMS and MCA. Thus, the storage limits are not fixed at the end of the month, and are fixed at the end of the study. Meanwhile, PCN and REV are assumed to have unchanged or “horizontal” forebay trajectories over all of the time intervals. Their forebay elevations are set to a higher value. This assumption is rational, in fact, and is verified by the results of the model with the water head convergence scheme.

In addition, it can be clearly realized that the forebay trajectory is indeed controlled by the elevation range at the end of each month. The sign (+) and (-) represent the relaxation limits of the range, allowing reservoir forebay to change at each month. Moreover, the range limits are scheduled in advance, usually from 1 to several feet, depending on the specific study requirements. In general, GOM studies usually require the reservoir forebay to end at a given level in the final time-step of the study. For other time intervals, the reservoir elevation is allowed to vary by the optimization algorithm within specified limits. For consistency, GOMMS will adopt the same settings as used in GOM.

![GMS Forebay Trajectory](image)

**Figure 4-6: GMS forebay trajectory and monthly limitations in 1964**
Figure 4-7: MCA forebay trajectory and monthly limitations in 1964

Figure 4-8: ARD forebay trajectory and monthly limitations in 1964
4.1.3. System Resource Balance and Markets Activities

Figure 4-9 illustrates the total system capacity, system load, reserve and residual system resources and imports/expert for the study load year 2008-2009. This figure is important as it clearly illustrates that during the spring-summer period, the load is low while the system capacity is high, which makes this period the best time for carrying out maintenance activities. Additionally, August and September are known to be unsuitable for maintenance activities because of the good export market opportunities during these months. At that time, the US market, in particular, South California has a heavy electric load, which increases the spot prices during these months, compared to other months (see Figure 4-10). Figure 4-10 also shows that early-October is an opportune time for maintenance, but with the start of winter in BC in late-October, only the first two or three weeks are considered to be suitable for maintenance scheduling. In any case, this period is still considered to be appropriate for maintaining units with S/C capability, since the system has enough rotational energy and the system load is still relatively low.

Figure 4-9: Load and resource balance plot (2008-2009 forecast load year)

Figure 4-10 illustrates the weekly average price forecasts in both the Alberta and the US markets. Clearly, the electricity market prices are high in the winter and in the summer months and low in other months, and the commercial strategies of BC Hydro are apparent for the markets. That is, import from Alberta and export to the US, and export more during those months when the prices are higher (i.e., winter and summer) and import at other times.
4.2. Numerical Studies and Comparisons

In this section, the models that were presented in Chapter 3 will be verified, analyzed, and summarized, one by one. The models are: the mixed-integer programming model without water head convergence based on formulas (3.7) - (3.12); the mixed-integer programming model with water head convergence based on formulas (3.13) - (3.30); and the simple linear relaxation scheme based on formulas (3.45) - (3.50). The numerical studies are done in terms of weekly time steps, but each model is verified by the schemes of both single, sub-time steps, and multiple, sub-time steps. Single sub-time steps means that every physical quantity, such as load, water inflow, or price, is calculated by the simple arithmetic average of the corresponding week. Multiple, sub-time steps consider that each physical quantity, though still measured as an average, should reflect the different characters with respect to the time divisions, i.e., LLH and HHL (light load hours and high load hours). For example, it is more accurate and meaningful that the daily loads are averaged by daytime and night time hours, separately, rather than by 24-hours per day. The average measurement also implies that, in the multiple, sub-time steps model, each original physical quantity has been treated as a vector with values representing the corresponding average value in terms of its time division. The use of two sets of schemes is helpful in that we may more clearly know the efficiency of each model and the influencing factors on the objective values.

To make the optimized outage scheduling more rational, some implicit considerations are taken from expert experience:
1. The units' maintenance would be finished before the middle of July so as to avoid export seasons in the summer.

2. The unit number that could be maintained simultaneously in the same plant is restricted due to the requirement of available crews. This simple rule can also effectively ensure that the system would always have enough rotational energy. In the following numerical studies, only in GMS is it allowable for two units to be off-line at the same time and all other plants can have maintenance once for one unit.

3. The arrangement for maintaining S\C is special and requires a greater consideration, compared to the regular units. (This conclusion is verified by GOM studies, based on the formulas in Section 3.2.7.) To support enough voltage for importing electricity, running S\C is economical because it can absorb reactive power from the grids; however, the alternative timings for maintaining S\C units are greatly numbered. Traditionally, only early-autumn (October) and spring (usually before May) are suitable for the units’ maintenance since, during these periods, loads and prices are relatively low and the system has enough units to support rotational energy and capacity. Since people are aware of this situation, and also for the sake of simplicity and efficiency, the maintenance timings for S\C could be restrained into preferable durations, instead of using binary schemes (3.41, 3.42) to determine the optimal periods. This simple way can avoid the senseless outage schedules, but cannot evaluate the “costs” for those S\C outages. In the case studies, we would restrict GMS 1 and 2 to be maintained only in October and let the model choose the “right” timings for the remaining S\Cs (i.e., GMS 3 and MCA 1 & 2). Obviously, more unit combinations are possible than if they are just restricted into the early-spring and autumn, and the model’s efficiency can be fully tested.

4. From bihourly GOM studies, one can observe that the maintenance of REV should be specially arranged in certain periods because of the excessive local inflows. This phenomenon can be captured and explained by a bihourly model, but performing the weekly time-step models would produce unclear results. To test the models, this concern is not considered in the constraints (though it would easy to do, temporarily). At the end of the chapter, we present the results (and magnitudes) when the REV unit is taken out of service during May and June, according to one GOM study.
4.2.1. Single sub-time step results

4.2.1.1. Mixed integer programming model without head convergence scheme

Figure 4-12 shows the unit outage scheduling results from the mixed-integer model without the water head convergence scheme. Apparently, the outage season is concentrated on the moments when loads are relatively low and the reservoir is nearly empty.

First, the outage schedules in GMS are analyzed. Ignoring the fact that GMS unit 3 is S/C, this table can give interesting explanations for arranging different unit types in GMS. In March, the reservoir elevation reaches the lowest level in the year. Thus, from the angle of plant efficiency, the units having a comparatively higher efficiency in smaller discharge or lower water head regions, should be in service at that time, since the plant needs to run frequently at those head/discharge zones. Currently, GMS units 1 to 5 (type 1) have relatively high efficiency in lower discharge regions, but unit 9 and 10 (type 3) are more suitable for running in larger discharge regions. The performance of units 6 to 8 (type 2) is in the “middle” between type 1 and type 3. Hence, removing two units (type 1) in the early-spring seems to be an unwise decision because the plant must operate frequently with smaller discharges. Similarly, taking units 9 and 10 out of service in the early-summer will lead to a risk in losing efficiency because the reservoir is nearly impounded and has sufficient water supply for the turbines. As a compromise, which has been verified to be the best strategy for this case, one unit (of type 1) and another unit (of type 3) should be removed together, to avoid a loss in efficiency in the lower discharge region or to lose the chance to generate more electricity, if necessary. Taking two units off service (both type 2) is acceptable because the remaining units can cover all unfavourable situations.

MCA 1 and 2 are also S/C and can only be maintained in the spring. Even without special S/C timing constraints, the results look satisfactory. The maintenance schedules in REV also seem to be appropriate since they avoid the water seasons in May and June. The result is surprisingly good, considering that the weekly time-step model is believed to be too coarse to capture this phenomenon. Even so, this may still only be a coincidence if the other models fail to arrive at the same conclusion. Finally, plant ARD should take off units when its reservoir reaches the lowest level (Figure 4-8), and thus, this schedule seems to be reasonable.

Figure 4-11 shows the changing laws of objective values and water head errors with respect to the iterative runs. Water head error is defined as the difference between the current total water head and the water head in the last run. The results indicate that by simple water head iteration, both objective values and water head errors converge into one certain value (not zero) and oscillate around it thereafter.
The mixed-integer model is solvable, and the solver (i.e., CPLEX 10.1) adopts dual and primal simplex as the MIP algorithm. Experience reveals that at every run or iteration, the gap between the primal and the dual problem is usually less than 0.03% within one hour. Therefore, it is reasonable to imagine that real optimal solutions can be obtained if the solver is given a longer time-span. Sometimes, achieving real optimal solutions is unnecessary, since, after a certain duration, (e.g., one hour, in this case), the error gap is already sufficiently small to be acceptable and thus, will not affect the final solutions by much.

As discussed in Section 3.2.8, it is not always easy to have linear constraints to arrange the units' outages in a consecutive way, especially in cases where one plant has multiple unit types and maintenance timings. If a "broken" outage schedule occurs in one plant, and the "split" duration is less than three weeks, it can often be designed with a new arrangement. First of all, based on expert experience, several alternative "movements" can be enumerated and by adjusting them somewhat, the maintenance activities can be connected to one another without any "breaks." If the numbers of alternatives are too many to be determined manually, then people can make use of the GOM studies, or historical maintenance records, to select an appropriate one. Although this method may not produce the real "optimal" answer, it is simple to do and produces solutions that are meaningful. In fact, from Figure 4-12, the outage of MCA unit 4 is seen to be advanced one week ahead, on the basis of expert experience.
Mixed Integer Programming Model without Water Head Convergence (single subtime step)

<table>
<thead>
<tr>
<th>Facility</th>
<th>Unit</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMS</td>
<td>13-OCT</td>
<td>2-NOV</td>
<td>10-MAY</td>
<td>09-JUN</td>
<td>19-MAY</td>
<td>08-JUN</td>
<td>06-JUN</td>
<td>19-MAY</td>
<td>05-APR</td>
<td>18-MAY</td>
<td>08-JUN</td>
<td>12-NOV</td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td>10-JUN</td>
<td>01-JUN</td>
<td>30-JUN</td>
<td>31-JAN</td>
<td>02-JAN</td>
<td>03-JAN</td>
<td>04-JAN</td>
<td>03-JAN</td>
<td>04-MAY</td>
<td>05-MAY</td>
<td>02-JUN</td>
<td>01-JUN</td>
<td></td>
</tr>
<tr>
<td>MCA</td>
<td>31-MAR</td>
<td>20-APR</td>
<td>11-APR</td>
<td>10-APR</td>
<td>09-APR</td>
<td>08-APR</td>
<td>07-APR</td>
<td>06-APR</td>
<td>05-APR</td>
<td>04-APR</td>
<td>03-APR</td>
<td>02-APR</td>
<td></td>
</tr>
<tr>
<td>REV</td>
<td>03-MAR</td>
<td>20-APR</td>
<td>11-APR</td>
<td>10-APR</td>
<td>09-APR</td>
<td>08-APR</td>
<td>07-APR</td>
<td>06-APR</td>
<td>05-APR</td>
<td>04-APR</td>
<td>03-APR</td>
<td>02-APR</td>
<td></td>
</tr>
<tr>
<td>ARD</td>
<td>31-MAR</td>
<td>13-APR</td>
<td>14-APR</td>
<td>13-APR</td>
<td>14-APR</td>
<td>15-APR</td>
<td>16-APR</td>
<td>17-APR</td>
<td>18-APR</td>
<td>19-APR</td>
<td>20-APR</td>
<td>21-APR</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4-12: Mixed-integer model without water head convergence
4.2.1.2. Mixed integer programming model with head convergence scheme

Figure 4-16 is the outage schedules based on the mixed-integer model with water head convergence constraints. The water head convergence scheme means that the forebay/head is treated as a variable in the model, which can be formulized by Equations (3-13) - (3-30).

By carefully analyzing the outcomes, several conclusions are made. First, GMS has the same maintenance strategy as that determined by the previous model. Although not every unit has the same maintenance timing in both of the models, the combos, or more exactly the representative combos, at each time interval, are exactly the same in both models. PCN begins the unit outage one week later than its start in the former model, a small and acceptable difference. An interesting phenomenon is seen for the PCN forebay elevation (also for REV). As stated in Section 4.1.2, the previous model took a pre-specified horizontal forebay trajectory in PCN/REV, and after several iterations, one can see that the forebay elevation vibrated irregularly in an allowable range (blue line in Figure 4-13). In the model with water head convergence, however, its forebay was always at the maximum level (red line). Because all of the units in PCN have a better efficiency at a higher forebay level, it is reasonable to believe that the optimal strategy for PCN operation is to always keep the forebay elevation higher, which increases the plant efficiency.

![Figure 4-13: PCN forebay comparison in two models](image-url)
The greatest difference occurs in the REV schedules where the units are arranged to be off from April to June. This is not true in the ordinary situation, and two possible reasons may explain this. 1) The week time-step may be too coarse to “seize” the plentiful local inflows in May and June; or 2) a more significant reason may be that MCA is holding more water at the same time and mitigates the water release down to REV. Because MCA has a higher plant efficiency compared to REV, the model chooses to let MCA store more water and generate the power later. As shown in Figure 4-14, on May 5, the forebay by the water head convergence scheme is 0.75m higher than without head convergence. This value might explain why the maintenance in REV is delayed for about two weeks. In addition, Figure 4-10 clearly indicates that the prices from April 15 to June 15 are much cheaper than those at other times, which drives the system to import more from outside. These reasons may explain why ARD postpones the maintenance for one week.

![Figure 4-14: MCA forebay comparison in two models](image)

The water head convergence strategy (3.13) - (3.30) is not easily solved. A direct solution is still impractical, so the algorithm discussed in the final part of Section 3.2.4 is implemented. That is, the original model is separated into two parts: the first part gives the best “outage schedules,” and then the second part takes this information as a known condition, to converge the water head. By doing this, the model becomes manageable. From experience, by spending 220
minutes (3.6 hours), the difference between primal and dual problem can be reached within a 0.38% gap. Basically, the process is still quite time-consuming.

![Graph](image)

**Figure 4-15: The objective values and water head errors with respect to iteration**

Figure 4-15 describes the varying rules by iterating the model several times. Apparently, this algorithm lets the water head converge rapidly (three iterations are sufficient). Objective values can be used in a similar way with respect to iterations. The “eventual” objective value gains roughly two million in addition to the value in the previous model. This value may reveal how the magnitude of the water head convergence scheme affects the total system benefits.
**Mixed Integer Programming Model with Water Head Convergence (single subtime step)**

<table>
<thead>
<tr>
<th>Facility</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4-16: Mixed-integer model with water head convergence**
4.2.1.3. Simple Linear Relaxation Scheme

The last model is the simple linear relaxation scheme. As pointed out in Section 3.4, though this scheme cannot be proven to arrive at "real optimal" solutions, its high solving efficiency and reasonable results still make it attractive and competitive. The optimal schedules are listed in Figure 4-19.

In GMS, the maintenance activity is about to be brought forward two weeks ahead. Nevertheless, the rules of grouping the maintained units must still comply with the principle that the unit with the better performance at larger discharges should be paired with those units having a good performance at smaller discharges. As shown in Figure 4-17, with water-head convergence constraints, GMS's forebay is always higher than the forebay from simple linear relaxation. This may explain why the difference between the objective values can be as high as 20 million. Figure 4-18 gives the varying rules of objective values and total water head errors with respect to iterations. The water head error also cannot converge to zero and oscillate around 10m.

![Figure 4-17: GMS forebay by the simple linear relaxation scheme and MIP model](image)

PCN and MCA have the same outage planning as shown for the MIP model, though the schedule of REV is postponed until May. This result confirms the conclusion that the weekly
model is too insensitive and is incapable of measuring the influences by local inflows on maintenance.

The simple linear relaxation scheme is highly efficient compared to the above two integer models. Only one or two minutes are needed in weekly time-steps for each run and the results are acceptable, to a certain extent. The mis-scheduling of REV is a common problem in all of the weekly models, not just for this model. In any case, only this simple linear relaxation scheme is promising enough to handle daily or even bi-hourly time-steps, which would be sufficient to reflect REV local inflow conditions. Moreover, simple linear relaxation is also applicable when doing primarily maintenance studies without requiring a high precision. The number of additional integer cuts that must be added (Formulae 3.44, 3.45, and 3.48) depends on specific cases. From experience, GMS usually needs more cuts (usually 25-30), and the others need around 10-15 cuts.

Figure 4-18: The objective values and water head errors with respect to iteration
**Simple Linear Relaxation Model without Water Head Convergence (single subtime step)**

<table>
<thead>
<tr>
<th>Facility</th>
<th>Unit</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMS</td>
<td>01-OCT</td>
<td>1</td>
<td>21-OCT</td>
<td>14-APR</td>
<td>04-MAY</td>
<td>05-MAY</td>
<td>26-MAY</td>
<td>26-MAY</td>
<td>15-JUN</td>
<td>16-JUN</td>
<td>06-JUL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>21-OCT</td>
<td>14-APR</td>
<td>04-MAY</td>
<td>05-MAY</td>
<td>26-MAY</td>
<td>26-MAY</td>
<td>15-JUN</td>
<td>16-JUN</td>
<td>06-JUL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>01-OCT</td>
<td>14-APR</td>
<td>04-MAY</td>
<td>05-MAY</td>
<td>26-MAY</td>
<td>26-MAY</td>
<td>15-JUN</td>
<td>16-JUN</td>
<td>06-JUL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td></td>
<td>1</td>
<td>17-MAR</td>
<td>13-APR</td>
<td>11-MAY</td>
<td>08-JUN</td>
<td>06-JUL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>14-APR</td>
<td>11-MAY</td>
<td>08-JUN</td>
<td>06-JUL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>12-MAY</td>
<td>08-JUN</td>
<td>06-JUL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCA</td>
<td></td>
<td>1</td>
<td>31-MAR</td>
<td>20-APR</td>
<td>11-MAY</td>
<td>01-JUN</td>
<td>22-JUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>21-APR</td>
<td>11-MAY</td>
<td>01-JUN</td>
<td>22-JUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>12-MAY</td>
<td>01-JUN</td>
<td>22-JUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REV</td>
<td></td>
<td>1</td>
<td>21-OCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25-MAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>01-OCT</td>
<td></td>
<td></td>
<td>25-MAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARD</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4-19:** Simple linear relaxation model without water head convergence
4.2.2. Multiple sub-time step results

4.2.2.1. Mixed Integer Programming Model

The use of multiple sub-time-steps is only for comparison purposes. The design of the following case studies is intended to: 1) check the efficiencies of different models, and 2) address the concern of whether or not, after using sub-time-steps, the weekly model will be adequate to see how local inflows of REV will have an impact on the outage decisions. In the subsequent sections, the outage of GMS unit 3 will only be allowed to occur during early-March to middle-April, in view of its SiC functionality. This change also allows the maintenance activity to have a few weeks of “break” that would otherwise be impossible to manually schedule in a consecutive way. Consequently, we propose a possible “maintenance shape.”

The multiple sub-time-steps are prescribed as follows: Every workday is divided into four divisions, of 2, 2, 12, and 8 hours. Every holiday or weekend is separated into 2 divisions: one of 16 hours and the other of 8 hours.

First, we focus on the mixed-integer model without water head convergence (Figure4-20). The rule of grouping maintained units is kept the same as for single sub-time-step cases. The units with different operational properties are paired so as to avoid extremely unfavourable situations that can lead to the loss of plant efficiency. Because the units’ maintenance must begin from early-March, breaks of approximately 5-weeks must take place between the maintenance activities. As stated before, all mixed linear/linear models do not contain any general constraints for maintenance continuity, especially in GMS, due to its multiple unit types. Unlike the case where “breaks” of only one or two weeks exist between maintenance periods, which can be manually adjusted to guarantee the connectivity of maintenance activity, this situation is more complicated. Either the outage of unit 8 would be delayed or the outage of another unit could be shifted to an earlier time. Based on these considerations, a more rigorous approach is needed from a theoretical angle. Hence, the “Maintenance Shape” approach is designed.
Mixed Integer Programming Model without Water Head Convergence (multiple subtime steps)

<table>
<thead>
<tr>
<th>Facility</th>
<th>Unit</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMS</td>
<td></td>
<td>01-Oct</td>
<td>21-Oct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01</td>
<td>01-Oct</td>
<td>21-Oct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>03</td>
<td>23-Mar</td>
<td>19-May</td>
<td>06-Jun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td></td>
<td></td>
<td>10-Mar</td>
<td>06-Apr</td>
<td>04-May</td>
<td>01-Jun</td>
<td>30-Jun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>02</td>
<td></td>
<td></td>
<td>07-Apr</td>
<td>05-May</td>
<td>02-Jun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCA</td>
<td></td>
<td></td>
<td></td>
<td>31-Mar</td>
<td>20-Apr</td>
<td>11-May</td>
<td>01-Jun</td>
<td>22-Jun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>03</td>
<td></td>
<td></td>
<td>12-Mar</td>
<td>01-Jun</td>
<td>02-Jun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REV</td>
<td></td>
<td>21-Oct</td>
<td></td>
<td></td>
<td>14-Apr</td>
<td>04-May</td>
<td>15-Jun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01</td>
<td></td>
<td></td>
<td></td>
<td>05-May</td>
<td>25-May</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARD</td>
<td></td>
<td></td>
<td></td>
<td>23-Mar</td>
<td>06-Apr</td>
<td>07-Apr</td>
<td>20-Apr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4-20: Mixed-integer model without water head convergence
4.2.2.2. Maintenance Shape Approach

Maintenance shape can be defined as one vector; whose length is the maintenance duration and whose member represent the unit number of the unit being maintained (at least equal to one). Naturally, the maintenance shape could ensure the continuity requirement because each vector member may be thought of as one time interval (i.e., one week, in our case study) during the maintenance activity and at least one unit must be maintained during that interval. The maintenance duration is not a fixed value, and can be any value under the constraints of continuity and other requirements (see Table 4-4). Following, is an intuitive example for the maintenance shape.

Table 4-2: illustrative Case 1

| Week | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 |
|------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| U1   | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U2   | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U3   | 0  | 0  | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U4   | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U5   | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U6   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  |
| U7   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U8   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  |
| U9   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  |
| U10  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  |

Table 4-2 (case 1) gives one possible situation for GMS, where the maintenance activity begins from the 21st week and persists until the 37th week. In total, 17 weeks are needed for all unit maintenances (maintenance duration). In each week, no more than two units are allowed to be out of service. The bold number one values indicate that the unit is undergoing maintenance.

Table 4-3: illustrative Case 2

| Week | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 |
|------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| U1   | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U2   | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U3   | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U4   | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U5   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  |
| U6   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  |
| U7   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U8   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U9   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| U10  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |

Table 4-3 (case 2) illustrates another situation where the maintenance ordering of unit 3 and 10 are switched. Still, the unchanged quantities are the total length of outage duration and the number of units being maintained in each week. In the past, we believed that the different types of units (3 and 10), and cases (1 and
2), were much different, but based on the definition of Maintenance Shape, these two cases are identical. Figure 4-21 plots the "definition" for the "shape" of units' maintenance.

![Figure 4-21: the "Shape" of maintenance activity](image)

The underlying reason for introducing maintenance shape roots was in the attempt to shrink the number of possible outage combinations. At present, maintenance shape captures the most significant or basic characteristics of maintenance activity, i.e., unit numbers that are maintained at each time interval. With regards to the problem of different unit type combinations, it can be completely handled by the constraints (3.32 - 3.34). After the simplification, all satisfied maintenance shapes can be enumerated by an outside program (for this case study, it can be written by FORTRAN). The results show that the GMS plant has only 1,416 possible shapes for maintenance satisfying all requirements and constraints. (After requiring that the outage of GMS unit 3 is done before April 15th, the difficulty of enumerating all possible "shapes" was effectively lowered.) This relatively small number allows the problem to be solvable.

As explained before, maintenance duration is not a fixed value. In this case study, all possible maintenance lengths for the GMS plant can be easily calculated. The lengths will vary from 12 weeks to 18 weeks. Table 4-4 gives the shape numbers under the different duration lengths of maintenance activity.

<table>
<thead>
<tr>
<th>Duration Length</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Shapes</td>
<td>1</td>
<td>10</td>
<td>50</td>
<td>175</td>
<td>325</td>
<td>428</td>
<td>427</td>
</tr>
</tbody>
</table>
Next, we must ensure that regardless of how the units are arranged and combined, the shapes of maintenance activity must be controllable. This simplification is already a useful part of the overall progress, but it is still impracticable for linear expressions because the varying start dates and maintenance lengths constitute a non-linear relation. To degrade this non-linearity into linear expressions, one of the variables must be eliminated. Thus, a special algorithm is designed. The main idea is to separate the common effects of start date and maintenance length, and solve the maintenance scheduling problem under a fixed start date and duration, at each time.

**Step 1:** Start from the first week, which is possible for the maintenance activities (for example from the beginning of March, i.e., the 23rd week);

**Step 2:** Start from the shortest duration of maintenance (i.e., 12 weeks, in the GMS case);

**Step 3:** Add new constraints to the original models:

\[
\sum_{s} Y_{s} \cdot \text{Shape}_{s,t-\text{Starting date}+1} = \sum_{u} \sum_{nt} \text{XTiming}_{f,u,nt} \cdot (1 - \text{ONES}_{f,u,nt,t})
\]

\[
\sum_{s} Y = 1
\]

In (4.1), parameter ONES comes from Formula (3.46), which designates zeros as the maintenance weeks and ones to other weeks. The binary variable, \(Y\), is newly introduced, with subscript \(s\) meaning the number of possible shapes under the current maintenance length. These numbers are listed in Table 4-4. The underlying reason for introducing maintenance shape roots was in the attempt to shrink the number of possible outage combinations. At present, maintenance shape captures the most significant or basic characteristics of maintenance activity, i.e., unit numbers that are maintained at each time interval. With regards to the problem of different unit type combinations, it can be completely handled by the constraints (3.32 - 3.34). After the simplification, all satisfied maintenance shapes can be enumerated by an outside program (for this case study, it can be written by FORTRAN). The results show that the GMS plant has only 1,416 possible shapes for maintenance satisfying all requirements and constraints. (After requiring that the outage of GMS unit 3 is done before April 15th, the difficulty of enumerating all possible “shapes” was effectively lowered.) This relatively small number allows the problem to be solvable.

As explained before, maintenance duration is not a fixed value. In this case study, all possible maintenance lengths for the GMS plant can be easily calculated. The lengths will vary
from 12 weeks to 18 weeks. Table 4-4 gives the shape numbers under the different duration lengths of maintenance activity. Of course, only one of these will be chosen, as represented by (4.2). Parameter Shape is achieved by a FORTRAN enumeration program that gives all the “shapes” with the format as listed in table 4-2 or 4-3. Constraint (4.1) requires the maintenance activity to satisfy one of the shapes given in a consecutive manner.

**Step 4:** Let the current maintenance length + 1. Then go to step 3 again and calculate the new optimal value. If the current length is already the longest, then go to step 5.

**Step 5:** Let the current starting date + 1, then go to step 2. If it is already the last week when the maintenance activity can take place, then stop.

As verified by numerical studies, the new constraints (4.1, 4.2) will not increase the computational burden of the original models. By considering the water head renewal process, the above algorithm will be very time-consuming. Therefore, step 3 may only permit one or two iterations for water head updates. This may bring a few errors into the results, but it remains as a worthwhile process, from a practical view. The original scheduling (figure 4-20) suggests that unit 3 would begin maintenance at the beginning of March, and following this requirement, Figure 4-22 gives the seven optimal values for seven different outage durations, beginning at the same time. Interestingly, the 15-week outage length looks to be the best among all of them.

**Figure 4-22: Optimal values at different outage durations starting from March**

According to this outcome, the new outage schedule, in a consecutive way, can be re-plotted in figure 4-24. Two features are noticed. One is the basic rule of grouping maintained
units to follow the same principle. This result is important and useful since, under every circumstance, type 1 and type 3 must always be grouped together. Additionally, due to the change of GMS outage, REV will advance its maintenance for six weeks. The other feature is the optimal value that is obtained by the multiple sub-time-steps model, which is a little smaller than the value obtained from the single sub-time-step model. This may be caused by the fact that unit 3 (S\C) is forced to be maintained in the early-spring. From Figure 4-23, this fact is obvious.

From another aspect, the multiple sub-time-steps model has similar changing rules as in the single sub-time-step model, when the water head is updated in an iterative way. The results will converge to become a relatively small error or quantity, but will always oscillate around that value.

Generally speaking, the computation for the multiple sub-time-steps model is more difficult than that for the single sub-time-step model. For the same time-span, the multiple sub-time scenario is roughly 1/8 - 1/10 as efficient, compared to the single sub-time-step. The results are still acceptable since, in a one-hour span, for example, the gap between primal and dual problems can be degraded into a 0.3% range (compared to a 0.03% range in the single sub-time-step). Therefore, it is rational to believe that, given longer time-spans, a greater precision can be achieved.

Unfortunately, for the water head convergence model, the computation cannot be completed in a reasonable length of time. Numerical studies show that a dual-primal algorithm will take 12 hours to degrade the gap into a 5% range. This larger error and the intolerable process would make the use of the water head convergence senseless. Even though the final result (with a 5% gap) is not optimal, the outage scheduling will be the same as in figure 4-20.
Mixed Integer Programming Model without Water Head Convergence (multiple subtime steps)

<table>
<thead>
<tr>
<th>Facility</th>
<th>Unit</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMS</td>
<td>1</td>
<td>01-OCT</td>
<td>21-OCT</td>
<td>03-MAR</td>
<td>23-MAR</td>
<td>04-MAY</td>
<td>05-MAY</td>
<td>25-MAY</td>
<td>15-JUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>01-OCT</td>
<td>21-OCT</td>
<td>24-MAR</td>
<td>13-APR</td>
<td>25-MAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>01-OCT</td>
<td>21-OCT</td>
<td>24-MAR</td>
<td>13-APR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>04-MAY</td>
<td>05-MAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>25-MAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td>1</td>
<td>10-MAR</td>
<td>06-APR</td>
<td>04-MAY</td>
<td>01-JUN</td>
<td>30-JUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCA</td>
<td>1</td>
<td>31-MAR</td>
<td>20-APR</td>
<td>11-MAY</td>
<td>01-JUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REV</td>
<td>1</td>
<td>21-OCT</td>
<td>10-MAR</td>
<td>30-MAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>01-OCT</td>
<td>10-MAR</td>
<td>31-MAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>20-APR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>11-MAY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARD</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4-24: Renewed maintenance scheduling in the consecutive manner
4.2.2.3. Simple Linear Relaxation Scheme

Finally, the simple linear relaxation scheme gives the optimal scheduling (figure 4-26). The differences between figure 4-26 and figure 4-24 are mainly present at two places: 1) GMS delays the maintenance for roughly three weeks; and 2) REV has different outage timing.

Although the simple linear relaxation scheme is efficient, the outage of REV seems to be barely acceptable. As discussed at the end of the last section, this is not the fault of the simple linear relaxation scheme since the weekly time step is too coarse to appropriate this phenomenon. Figure 4-25, which is from a bihourly GOM study, gives the potential loss if one unit is taken out for one week during the whole year. This result is obtained from other prices and loads in forecasting cases, and is based on averages from 10 water years (from 1964 to 1973). Hence, the value may not be in consistent with this case study. Nevertheless, this plot intuitively suggests that taking a unit out in May and June is even worse than taking a unit out in the winter.

![REV Unit Outage Costs](image)

**Figure 4-25: REV one unit outage cost**
Simple Linear Relaxation Model with Water Head Convergence (multiple subtime steps)

<table>
<thead>
<tr>
<th>Facility</th>
<th>Unit</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMS</td>
<td>01-OCT 1</td>
<td>21-OCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01-OCT 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCA</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REV</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARD</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4-26: Simple linear relaxation model without water head convergence
4.3. Other Measures Concerning Solving Efficiency

CPLEX is a well-known linear programming solver with the most efficient solving process. When used for integer programming, CPLEX is reported to be 10 to 100 times faster than other commercial programs. By 2006, ALOG Company released its latest CPLEX version 10.1 with great enhancements in the integer programming, in contrast to the earlier version 9.0. Nevertheless, the MS problem is essentially a \( \mathcal{NP} \)-hard problem that would be difficult for any available algorithm. (For a brief discussion of its properties, refer to Appendix D.) Thus, even the recently developed algorithms may not be useful for MS problems. From practical experience, with certain constraints involving more binary variables, no algorithm is completely effective in producing optimal solutions. In addition, CPLEX provides many algorithm alternatives for users, such as primal simplex, dual simplex, and barrier with crossover, etc. Each of these could be more adaptive and effective in some aspects. Also, the multiple choices can be confusing to users who may be unsure of which algorithm best suits their case. Some slight modifications in the formulation may greatly change the behaviors of certain algorithms without having an impact on the final results. From these facts, it is meaningful to have a brief discussion on the solver issues, including the most recommended algorithm and advice or suggestions about formula transformations. With this knowledge, users could be more aware of the most important properties inherent to integer models, such as reasonable formula structures and appropriate algorithms, with their sensitively that can affect the solving efficiency.

In CPLEX 9 or the earlier version, barrier algorithm with crossover is usually used as the default algorithm. With very slight changes to Formula (3.33) and (3.34), however, the solving efficiency can be enhanced by 10 times. The possible explanations for this are as follows:

In combinatorial theories, constraints (3.33), (3.34) are called Generalized Upper-Bound Constraints, which have a standard form like:

\[
\sum_{j \in Q} x_j = 1
\]

(4.3)

Although, conceptually, this type of constraint is right and tightening, their solution needs more attempts. Commonly, commercial IP codes, such as the famous Branch-and-Bound algorithm, are usually associated with LP relaxation. The general procedure is to solve LP relaxation first, then look at the value of each variable. If it is integer, then keep it, otherwise, a routine called "dichotomy" is employed which expands an enumeration tree for searching for an integer. This selection routine is too complex for this discussion. Besides the Generalized Upper-Bound Constraints, the more common constraints for the LP model have the following structure:
\[ \sum_{j \in Q} a_j x_j \leq b_j \text{ or } A \cdot x \leq b \]  \hspace{1cm} (4.4)

These two constraints are treated in MIP or BIP algorithms, and have a totally different dichotomy. Figure 4-27 shows their branching schemes.

\[ \sum_{j \in Q_1} x_j = 0 \quad \sum_{j \in Q_1^c} x_j = 0 \quad x_j \leq \lfloor x_j^0 \rfloor \quad x_j \geq \lfloor x_j^0 \rfloor + 1 \]

(a) \hspace{2cm} (b)

**Figure 4-27: Division process of enumeration trees**

The left-hand plot (a) explores how variables in (4.3) are branching. Here, \( Q \) is a natural number set; with \( Q \subseteq N \). \( Q_1 \) is a nonempty subset of \( Q \). Now, looking at constraint (3.33), for example, for plant GMS at every time-step, if a total of 10 combos are available, then the binary variables \( g \) will have 10 candidates, i.e., \( Q = \{1, 2, \ldots, 10\} \). On the right-hand plot (b), square brackets denote the function of rounding the real value into the close integer.

The usual step, using LP relaxation to solve the IP problem, is to implement the division at each node by adding linear constraints, for example, in the union of constraint sets \( S = S^1 \cup S^2 \), with \( S^1 = S \cap \{ x \in \mathbb{R}^n_+ : d \cdot x \leq d_0 \} \) and \( S^2 = S \cap \{ x \in \mathbb{R}^n_+ : d \cdot x \geq d_0 + 1 \} \) where \( (d, d_0) \in \mathbb{Z}^{n+1} \) satisfies \( d_0 < d \cdot x < d_0 + 1 \).

Hence, for the Generalized Upper-Bound Constraints, the selection of \( d \) and \( d_0 \) is based on subset \( Q_1 \). The choice of this \( Q_1 \), however, is not easy. Still using the example from GMS, a desirable way to do this is to divide \( Q \) into the subsets \( Q_1 \) and \( Q \setminus Q_1 \) of roughly equal size. Assuming that \( Q_1 = \{1, 2, 3, 4, 5\} \), then its supplementary set \( Q \setminus Q_1 \) will be \( \{6, 7, 8, 9, 10\} \), as shown in Figure 4-27. One of the equations beside the branches will be added into the original model. As this searching routine goes deeper, more of the “constraint cuts” will be generated so that the model size will become correspondingly larger.

In a different way, if the model only has constraints like (4.4), the selection of \( d \) and \( d_0 \) will not increase the “cuts” to the original model. In recalling the procedure of the Branch-and-Bound algorithm, the newly added branch only adjusts or updates the lower and upper bounds of the existing constraints. This division brings an important advantage with it, in that the size of
the basis will never increase (Nemhauser, 1988). Thus, its solving efficiency should be at least as good as that of the Generalized Upper-Bound Constraints.

Even without knowing the exactly technical details of how the CPLEX 10.1 works, the underlying mathematical principle does not vary. Thus, it is reasonable to believe that by relaxing (3.33) - (3.34) into the forms of (4.5) - (4.6), the computation efficiency will be further promoted.

\[ \sum_{m} g_{j,m,t} \leq 1 \quad (4.5) \]
\[ \sum_{m} XT \text{Timing}_{j,u,m,t} \leq 1 \quad (4.6) \]

**Proposition:** If plant’ generation has minimum requirement larger than zero, then by changing constraints (3.33-3.34) into (4.5-4.6), the optimality of the MS model will not change.

**Proof:** Designate set \( S \) as constraints (3.32) - (3.34) and \( S^l \) as constraints (4.5) - (4.6). Obviously, we have \( S \subseteq S^l \). So, the optimal value \( V^l \) on \( S^l \), is at least as good as the value on set \( S \) (i.e., \( V^l \geq V \)). Now, suppose that a case exists with set \( S^l \setminus S \), and the optimal value \( V' \) is larger than \( V \), namely, \( V' > V \). This means that at least one of the variables, \( g \) or \( XT \text{Timing} \) must satisfy the constraint \( \sum x_j = 0 \) or \( x_j = 0, \forall j \). Regardless of which variable have all zero values, by (3.32) we know that combo designator \( g \) cannot be any other value but zero. From discharge restriction (3.11), no discharge will be released. Then, the output should also be zero, which is against the assumption of the proposition. Therefore, the situation of either \( g \) or \( XT \text{Timing} \) all being zeros is not possible, and then, \( S = S^l \) and \( V^l = V \).

This proposition tells us that by replacing the equality into inequality, the final optimal value will not be influenced. Although, in the current model, not all plants have minimum output requirements, this can easily be handled by setting a small positive value, say 0.1 MW. Also, the optimal maintenance scheduling by (3.32) - (3.34) does not have to be the same as those in (4.5) - (4.6), because of the different solving strategies. The evidence, from several practical studies, shows that this modification can promote the efficiency by roughly 10–times, compared to the original model using CPLEX 9.1.

For the latest version of CPLEX 10, the most efficient algorithm for the MS case is the dual and primal simplex approach. It is roughly twice as fast as the barrier algorithm of CPLEX 9.1. Regardless of how the constraints are formed, either (3.33, 3.34) or (4.5, 4.6), the solving
efficiency is virtually unchanged. This is true even with equations 3.33 and 3.34, which is slightly faster than equations 4.5 and 4.6. Thus, we conclude that the MS problem is indeed algorithm-specific, and different potential algorithms and formulation styles should be evaluated for users to arrive at the best efficiency.
CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS

5.1. Summary

This thesis focused on the potential use of mixed-integer/linear programming algorithms to solve the maintenance scheduling problem for a large scale hydroelectric system. The objective was to determine the best “timing” for each unit outage in the system. As the unit maintenance problems play a major in hydro system operation, one simple idea that could be followed was to use an existing hydro scheduling model and add a new set of constraints that satisfy all the existing power and non-power requirements and that could potentially derive the “optimal” maintenance schedules for each unit included in a study. This thesis followed this idea by firstly introducing and reformulating all the equations in the hydro scheduling model that may affect the maintenance decisions and secondly by concentrating on the constraints that could effectively satisfy the maintenance scheduling requirements.

Integer programming methods are usually considered suitable for maintenance scheduling problems. But in view of the complexities inherent in hydro models and in mixed-integer programming algorithms, it is very difficult to generalize one standardized approach to effectively handle this complex and large scale problem. Therefore, this research has focused on two main points. One focused on decreasing the number of possible binary variables in the model. The other was to find an appropriate algorithm that could approximate and reasonably simplify the computational process and to transform the nonlinear constraints into linear ones. It was also concluded from the literature that without specifically or delicately designing the maintenance scheduling constraints, the problem will be hardly solvable even by the latest commercially available algorithms.

After completing the discussion on problem formulation, a simple case study was carried out. Although this case study was too simple to be used as a guide for modeling the real maintenance scheduling BC hydro system, it however reveals some interesting properties of the integer programming model as detailed below.

- A mixed-integer maintenance problem belongs to NP-hard problems that could not be solved by polynomial time algorithms.
- The most appropriate mixed-integer maintenance scheduling model would be formulated in such a way that it develops a maintenance schedule that ensures that the plants are operated in to maximize system efficiency.
• Head convergence has an influence on hydroelectric unit outage scheduling algorithm. Head convergence will slightly impact the water balance calculations in the model and will therefore impact the efficiency of the up- and downstream plants in a river system. Modelling variations in head is not an easy process however and it would impact the solution efficiency of the maintenance scheduling model. Therefore the recommended algorithm in this research does not incorporate modeling variations in head for hydroelectric systems.

• The current version of the maintenance model doesn’t incorporate modeling of the S\C state.

• The simple linear relaxation scheme is very efficient as compared to the integer model, although there are no guarantees that the current implementation can yield the real “optimality”, the results seems to be acceptable in practice.

• The model does not contain equations to handle continuity of the outage schedules for a particular plant. To handle this problem a set of rules were developed to ensure that no gaps are present in the outage schedules, in particular if these gaps are one or two weeks long. If the gaps are too long, the proposed “maintenance shape” method may be helpful although it usually takes longer time to complete.

• In the case studies the rotational energy requirement was mainly satisfied in large part because the restrictions on the allowable number of outage units for a given plant.

• The case studies show that two units capable of running on S\C can usually be taken out for maintenance in the month of October at the GMS plant. Other units can be scheduled in the period from early march to the end of June. It is interesting to note from the results of the case studies that units type 1 are grouped together with units’ type 3 in the maintenance schedules. And it seems that pairing two units of different types is a preferable plan than performing the unit maintenance for individual units one at a time.

• The case studies have indicated that the maintenance schedules for the PCN and the MCA plants are not very sensitive, as the outages maintenance schedules did not change in the different model formulations that have been attempted. However, it is believed that this is not the general case; in particular when other factors are changed. For example, due to the restrictions or limitations on MCA’s forebay, the maintenance schedules for its units usually do change.

• It was found that the REV plant maintenance schedules couldn’t be accurately derived using weekly time step models for the months of May and June due to high natural inflows during this period. The case studies have proved that this period is not the appropriate to carry out maintenance for the REV units. However multiple sub-time steps model could results in better
optimal results than single time step model, although it usually could take longer time to solve the problem.

- The ARD plant is small and the maintenance of its units is usually carried out when the reservoir is at low storage level in the periods starting at the end of March and early April.

5.2. Future Development

As stated at the beginning of the first paragraphs in this chapter, the main purpose of this thesis was to use a deterministic integer programming techniques to solve the maintenance scheduling problem for hydroelectric systems. In reality, there are many uncertainties in the hydro system and in electricity markets. Therefore the technique developed in this thesis could potentially be enhanced to address such uncertainty if it is to be used in practice.

Future research on the maintenance scheduling problem could mainly concentrate on the stochastic properties of the hydroelectric generating system, and the research scope could be extended to cover all hydro and thermal units if needed. But because of the complexities and uncertainties of the hydro system, the problem scale may become too large and uncontrollable as multiple types of constraints are added, which make the problem more realistic and closer what is confronted in practice, and as the number of time steps increases. Moreover, the maintenance scheduling problem is essentially an integer programming problem that is much harder to solve than linear programming problems. None of these difficulties are easy to be conquered, so future research should focus on finding an appropriate modeling methodology and "ideas" to strike a reasonable way to overcome these difficulties given the current and future computation possibilities. Most importantly, the methods and techniques developed should focus on generating convincing schedules that could be implemented in practice. The followings are some potential future developments to the work carried out in this research.

- Development of practical scheduling methods. Scheduling methods are usually based on certain criterion or priorities. For example, the maintenance cost for each unit at certain moment in time is commonly used in practice. By rationally simplifying the problem the problem will become controllable and solvable. These methods can also take care of some of the uncertainties in this complex problem, but the problem is that the results will be dependant on the initial assumptions and simplifications used, and could therefore lead to sub-optimal schedules.

- Stochastic linear programming methods. The Stochastic linear programming technique is rooted in deterministic linear programming. By introducing the scenarios, stochastic linear
programming models can handle uncertainties in some ways. Nevertheless, the stochastic linear programming technique becomes unsuitable for large scale problems. In addition, when integer variables are introduced to properly model maintenance scheduling problem they become unsolvable.

- Stochastic dynamic programming methods. Dynamic programming is a natural technique to solve integer problems. The problem, however, is that the dynamic programming problem dimensions increase considerably as the number of state variables increases and the problem become impossible to solve for traditional computers. Certainly, advanced techniques like parallel computations and decomposition techniques will help in solving the problem, but the cost and time needed to obtain a solution becomes un-negligible.

- Real options or decision trees. These methods are similar in certain ways as they all try to answer the questions of "what-if". By calculating all the "costs" at each decision node, it will be easy to determine how good each decision is and to calculate the expected cost of all possible alternatives. But the difficulty may still come from the "dimensionality expansion" on the condition that the system has a set of "states". From this point of view, these have similar properties to dynamic programming methods.

Finally, it could be concluded that there is no one method that can easily handle the stochastic maintenance scheduling problem for hydro systems, and this makes this topic very interesting and attractive for future research. Probably the future methodology will be based on a combination of several approaches. No matter what the method will turn out to be, it will be definitely full of new challenges and opportunities.
BIBLIOGRAPHY


Chang, G. W.; Su, C. T. “A Practical Mixed integer Linear Programming-Based Short-Term Hydro Scheduling”, *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES*


Gonzalez, Camino; Juan, Jesus; Mira, Jose; Francisco J. Prieto and Maria J. Sanchez “Reliability Analysis for System With Large Hydro Resources in a Deregulated Electric Power Market”, *IEEE transactions on Power System*, Vol.20, No.1, Feb. 2005


Trick, Michael A. “A Linear Relaxation Heuristic For The generalized Assignment Problem”, 1991, working paper


APPENDICES

Appendix A. Linear Interpolation Scheme

This part will give two propositions indicating the interpolation properties by equations (3-13) to (3-30). Basically 1D approximation scheme (3-15, 16 and 17) is linear interpolation. And 2D case by (3-13-30) has very good approximation effects, with error bounds being formulaically evaluated.

**Proposition A.1:** If equation (3.15) holds, then approximation by formulas (3.16) and (3.17) are linear interpolation with regard to forebay.

![Figure Appendix A-1: Linear interpolation of (3-16) to (3-17)](image)

**Proof:** According to (3-21), (3-22) and (3-25), we have: \( \lambda_{j_b} + \lambda_{j_b+1} = 1 \).

Because (3-15) holds, e.g. \( FB_0 = FB_{j_b} \cdot \lambda_{j_b} + FB_{j_b+1} \cdot \lambda_{j_b+1} \), so linear relation:

\[
\bar{P}_0 = \bar{P}_{j_b} \cdot \lambda_{j_b} + \bar{P}_{j_b+1} \cdot \lambda_{j_b+1}
\]
will always be hold. Therefore, \( \bar{P}_0 \) is a linear approximation of \( \bar{P} \).

**Proposition A.2:** If equations (3-14) to (3-15) hold, then formula (3-13) associated with (3-20) to (3-26) satisfies the approximation property.

**Proof:** A little different from the first proposition, although formulas (3-20) to (3-26) are also linear forms, the approximation is not in the sense of linear interpolation. Let’s take illustrative figure (A-2) showing one slice of figure (3-2). For simplicity reason, the variable notations, as shown in the plot, are not totally identical with those in (3-20) to (3-26). Let’s suppose the
weighting factor $\theta$ in (3-13) to (3-15), which essentially is a vector, has four values at the corners of target rectangular grid which are presented by $\alpha, \beta, \gamma$ and $\rho$ (at least one of them should be nonzero).

Figure Appendix A-2: Approximation in generation surfaces (eqn. 3-13)

Firstly by (3-26), we have:

$$\alpha + \beta + \gamma + \rho = 1 \quad (A.1)$$

Then, by (3-15) and (3.2.3.17'), there is

$$f_{b_0} = f_{b_1} \cdot (\alpha + \beta) + f_{b_2} \cdot (\gamma + \rho) \quad (A.1)$$

Now we want to determine the difference between the real output value $P_0$ by linear interpolation and the approximated power value $P_0'$ by (3-13).

Due to (3-13), we have

$$P_0' = P_2^n \cdot \alpha + P_2^{n+1} \cdot \beta + P_1^n \cdot \rho + P_1^{n+1} \cdot \gamma \quad (A.2)$$

According to piece-wise linear relations between forebay and power output, we have:

$$P_2^n = f_{b_2} \cdot m_n + c_n \quad (A.3)$$

Index $n$ is the segment of the piece-wise linear function. Similar equations can be easily applied to the other three $P$ values. Substitute (A.3) etc into (A.2), then,

$$P_0' = c_n (\alpha + \rho) + c_{n+1} (\beta + \gamma) + f_{b_1} (m_n \alpha + m_{n+1} \beta) + f_{b_2} (m_n \rho + m_{n+1} \gamma) \quad (A.4)$$

From figure A-2, easily we can determine the linear function of $L$ (dot line) because it passes through the real value $P_0$.

The slope of linear function $L$ is:

$$P_0'' = f_{b_0} \cdot m_n + c_n \quad (A.5)$$
Because (3-14) is hold as an assumption, so
\[ Q_0 = Q_n \cdot (\alpha + \rho) + Q_{n+1} \cdot (\beta + \gamma) \]  
\[ (A.8) \]

Using linear function \( L \), it is straight to determine that \( P_0 = k \cdot (Q_0 - Q_n) + P_0^n \). Substitute (A.8) into it and simplify the formula, we have:
\[ P_0 = c_n (\alpha + \rho) + c_{n+1} (\beta + \gamma) \]
\[ + f_{b_1} [m_n (\alpha + \beta) (\alpha + \rho) + m_{n+1} (\alpha + \beta) (\beta + \gamma)] \]
\[ + f_{b_2} [m_n (\rho + \gamma) (\beta + \gamma) + m_{n+1} (\rho + \gamma) (\alpha + \rho)] \]
\[ (A.9) \]

Continue simplifying the expression,
\[ P_0 = c_n (\alpha + \rho) + c_{n+1} (\beta + \gamma) \]
\[ + f_{b_1} \left[ m_n \alpha \cdot \left( 1 + \frac{\beta \rho}{\alpha} - \gamma \right) + m_{n+1} \beta \cdot \left( 1 + \frac{\alpha \gamma}{\beta} - \rho \right) \right] \]
\[ + f_{b_2} \left[ m_n \gamma \cdot \left( 1 + \frac{\beta \rho}{\gamma} - \alpha \right) + m_{n+1} \rho \cdot \left( 1 + \frac{\alpha \gamma}{\rho} - \beta \right) \right] \]
\[ (A.10) \]

Then, calculate the difference between the approximation \( P'_0 \) and real value \( P_0 \):
\[ \Delta = P_0 - P'_0 = 
\]
\[ f_{b_1} [m_n (\beta \rho - \alpha \gamma) + m_{n+1} (\alpha \gamma - \beta \rho)] + f_{b_2} [m_n (\beta \rho - \gamma \alpha) + m_{n+1} (\alpha \gamma - \rho \beta)] \]
\[ (A.11) \]

Notate \( \epsilon = f_{b_2} - f_{b_1} \) and replace \( f_{b_2} \) in (A.11), then
\[ \Delta = P_0 - P'_0 = \epsilon (\beta \rho - \gamma \alpha) (m_{n+1} - m_n) \]
\[ (A.11) \]

Because \( \alpha + \beta + \gamma + \rho = 1 \) and all of them are nonnegative, so (A.11) can be reformed as
\[ |\Delta| = |P_0 - P'_0| \leq 0.25 \epsilon (m_{n+1} - m_n) \]
\[ (A.12) \]

Inequality (A.12) is very useful because it directly tells people the error boundary of this approximation. So, if people expect higher precision, one way is to use finer forebay increment (less…. value). The other way is to increase the segments of P-Q piece-wise linear functions, which can result in smaller slope changes between two adjoining segments. For example, now GMS is running under combo 1023 (10 units are all on-line), if the forebay increment is one meter and the interpolated point (\( P \)) is within the second segment, then this approximation will
Appendix B. General Equations for Unit Maintenance Decision

This set of formulas stem from Chang (2001, 2002). Basically they are believed to be general enough for any decision cases similar to unit commitment problem, such as maintenance scheduling problem. In BC Hydro system, each unit can be stipulated to have only two statuses, i.e., either on or off. So using a binary variable is proper enough to distinguish them. In section 3.2.4, one specially designed formula set is presented, associated with the characteristics of BC Hydro system aiming to accelerating the solving process. But it is very case-specific and maybe not suitable for any other systems.

There do be more units' conditions in other systems. For example, some units can work as pumps. In this case, pumping is another working status. So binaries must be replaced by integers such that additional status can be rightly represented. It is very necessary to have one set of formulas that can cover all kind of similar situations. Following formulas (B.1) to (B.8) are therefore developed in terms of such backgrounds and they can easily be extended to other cases by using integer variables. Although they have comparatively low efficiency in large scale problems, they are still worth being introduced in view of their generality.

\[
\sum_{m} g_{j,m,t} = 1 \tag{B.1}
\]

\[
g_{j,m,t} \cdot YC_{j,m,u} = U_{j,u,t} \tag{B.2}
\]

\[
U_{j,u,t} - U_{j,u,t+1} = US_{j,u,t} - UD_{j,u,t} \tag{B.3}
\]

\[
\sum_{t} US_{j,u,t} = 1 \tag{B.4}
\]

\[
\sum_{t} UD_{j,u,t} = 1 \tag{B.5}
\]

\[
\sum_{t} t \cdot (UD_{j,u,t} - US_{j,u,t}) = MD_{j,u} \tag{B.6}
\]

\[
\sum_{t} (t \cdot US_{j,u,t}) \geq MS_{j,u} \tag{B.7}
\]

\[
\sum_{t} (t \cdot UD_{j,u,t}) \leq ME_{j,u} \tag{B.8}
\]

Constraints (B.1) to (B.8) can determine the maintenance timing for each unit. Equation (B.1) means that only one combo could be chosen in each plant, at any time step. Combo,
invented in the GOM model, is a concept representing the units’ on-off combinatorial situation. Because GOM is using the efficiency curves at the plant level, so it is very needful to create a manner which can linearly connect unit combo and individual unit’s status. Currently (B.9) is the formula used to calculate the combo value in GOM studies. \( U \) is the binary number specifying unit’s on-off status, and subscript \( u \) is the index of unit. It is easy to verify that any unit’s combinatorial condition would have exclusive one combo value to match.

\[
\sum_{u} U_{j,u} \cdot 2^{(u-1)}
\]  

Equation (B.9) is critical because it links the relations between single unit’s status and every candidate unit combo. \( YC \) is binary parameter, specifying the unit on/off (1/0) condition in each combo. For example, GMS now is running under the combo 999 that means unit 4 and 5 are out of service. Hence, people could organize this information into parameter \( YC \) with following format:

\[
YC_{GMS,999,1} = 1, \ YC_{GMS,999,2} = 1, \ldots, \ YC_{GMS,999,4} = 0, \ YC_{GMS,999,5} = 0, \ YC_{GMS,999,6} = 1, \ldots, \ YC_{GMS,999,10} = 1.
\]

By this means, if the unit combinatorial condition is decided, it is straight to know the binary values from parameter \( YC \) by (B.9). However, in the case that plant has many units (e.g, GMS), listing all combos and their corresponding units status seem to be an arduous task. So it is preferable if a special program can be designed to for enumerating all combinatorial possibilities.

By this means, if the unit combinatorial condition is decided, it is straight to know the binary values from parameter \( YC \) by (B.9). However, in the case that plant has many units (e.g, GMS), listing all combos and their corresponding units status seem to be an arduous task. So it is preferable if a special program can be designed to for enumerating all combinatorial possibilities.

Equation (B.3) is created to guarantee the continuity of maintenance duration. Figure B-1 shows a simple example. Suppose that the 5th and 6th week will be selected for maintenance activity, hence \( U \) values should be all 1s except only those two time steps can have zero values. \( US \) is the variable specifying the start of maintenance activity, and similarly \( UD \) is used to record the end of the maintenance. If assuming that each unit would be taken out of service just only once, so either \( US \) or \( UD \) of each unit would have one time step to be 1 and all others are 0s.

Equation (B.6) indicates that the time steps between the start and end of maintenance activity must be equal to the pre-specified time duration. Constraint (B.8) and (B.9) are designed intentionally to eliminate those timings outside of the maintenance seasons. In BC hydro system,
wintertime, namely from November to next March, is usually not suitable for maintenance activities due to very high load demands. Therefore, most of scheduled maintenance will happen in the seasons of the spring and early summer (i.e., maintenance seasons), because at that time loads are not intense, market is less competitive and reservoirs are keeping impounding.

Appendix C. Simple Introduction on FCPA

Among linear relaxation schemes, the most widely-used method is so-called Fractional Cutting-Plane Algorithm (FCPA). Let's give some brief introductions on the main steps of this algorithm. Firstly donate F to the original constraint set, and π to constraint cuts/sets, which could be generated by certain rules (see the following discussions, e.g. strong valid inequalities), or by other ways. Suppose it is created by the valid inequalities, then step 3 describes the principles of generating those cuts. This step is also called as Separation Problem in mathematics. As an intuitive explanation, Separation Problem means at each run, only certain constrains from constraint set π will be added into the model. One of the underlying reasons is, it is very possible that set F is remarkably huge and lots of constraints in it are not “useful” or bounding to the solutions. Hence one effective way is to exclusively add the constraints from last run with the most “tighten” effects, or more strictly saying, add those constraints with the most violations against integrity into the current run, and then gradually force all the variables to be exclusive integers.

This algorithm is associated with a very widely used concept called strong-valid inequalities. As stated in step 4, constraint set F* is added to further restrict the model. Without being involved too many theoretical discussions, here only several useful notions are introduced as follows. Firstly standardize the constraint set S for 0-1 programming problems as the format:

\[ S = \{ x \in B^n : \sum_{j\in N} a_j x_j \leq b \} \]  \hspace{1cm} (C.1)

Then the class F of cover inequalities is given as:

\[ \sum_{j \in C} x_j \leq |C| - 1 \]  \hspace{1cm} (C.2)

Where C is a natural number, and C \subseteq N is called a cover if \( \sum_{j \in C} a_j \geq b \). Now,

The FCPA procedure with adding strong valid constraints is:

1) Initialization: \( S_R^t = \{ x \in R^n : Ax \leq b \} \), set iteration counter t to 1.

2) Solve the relaxation of LP(F), let \( x^f \) be an optimal solution. If they are satisfying integrality requirement, the stop, otherwise,
3) For given point \( x_j' \in R^n \setminus B^n \), find a \( C \) (assuming that one exists) with \( \sum_{j \in C} a_j \geq b \) and \( \sum_{j \in C} x_j' > |C| - 1 \). By doing this, introduce a vector \( z \in B^n \) to represent the unknown set \( C \), attempt to choose \( z \) such that \( \sum_{j \in C} a_j z_j > b \) and \( \sum_{j \in C} x_j' z_j > \sum_{j \in C} z_j - 1 \), or equally in the form \( \sum_{j \in C} (1 - x_j') \cdot z_j < 1 \). Then obtain the Separation Problem for Cover Inequalities:

\[
\xi = \min \left\{ \sum_{j \in \mathcal{N}} (1 - x_j') \cdot z_j : \sum_{j \in \mathcal{N}} a_j z_j \geq b + 1, \ z \in B^n \right\}
\]

Solve above model, if \( \xi \geq 1 \), then \( x_j' \) satisfies all the cover inequalities for \( S \). Otherwise \( \sum_{j \in C} x_j' \leq |C| - 1 \) is the most violated cover inequality for \( S \), add this constraint into additional constraint \( \pi \).

4) Refinement, Let \( \mathcal{F}' \subset \mathcal{F} \) be the set of one or more inequalities \((\pi, \pi_0)\) with \( \pi \pi' > \pi_0 \), and \( S_{R_k} = S_{R_k} \cap \{ x \in R^n : \pi x \leq \pi_0 \ \text{for} \ (\pi, \pi_0) \in \mathcal{F}' \} \)

5) Let \( t = t + 1 \)

In maintenance schedule case, the other constraint forms also exist, like equality (3.32). Those make things not as straight as premise assumed in FCPA procedure. Furthermore, as discussed before, the iteration is performed with the forebay updates, so it may bring unknown effects on the final optimality. Based on those factors, the rigor mathematical proof can barely be provided that using FCPA associated with strong valid constraints will make the solution approach the real optimality.

Let's simply explain how step 3 would work. Suppose unit 1 of GMS plant has 5 possible starting timing, after the first run, \( \text{XTiming}_{\text{GMS,1,nt}} \) (abbrev. by \( \hat{X} \)) gives the "split" results as: \{0, 0, 0.2, 0.3, 0.5\}. So, Separation Problem for Cover Inequalities is:

\[
\xi = \min \left\{ l \cdot z_1 + 1 \cdot z_2 + 0.8 \cdot z_3 + 0.7 \cdot z_4 + 0.5 \cdot z_5 \right\}
\]

s.t.:
\[
z_1 + z_2 + z_3 + z_4 + z_5 \geq 2, \ z \in B^5
\]

This simple problem is very obvious that the optimality \( \xi = 1.2 \) can be reached if and only if \( z_4 \) and \( z_5 \) are ones. According to the definition, current \( x_j' \) satisfies all the cover inequalities for \( S \) and no violated cover inequalities could be found. From above example, some facts can be summarized:
1) Essentially, the sense of cover inequalities could be explained in the way that the variables with biggest “split” values would have the strongest violations against the constraints.

2) MS problem couldn’t be found for suitable strong valid inequalities. The reason is, because the right-hand sides of (3.33) and (3.34) are always ones, so \( \zeta \) value will never be less than 1. So depending on step 3 to generate enough stronger constraint cuts seems impossible. In addition, even if step 3 can work after some modifications are implemented for MS constraints, it also induces another binary problem and, involves in too many subjective choices based on people’s judgements. Just like the comments in Nemhauser’s book, maybe the most effective way in terms of cover inequalities is by heuristics searching, but that is beyond the interests of this thesis.

**Appendix D. Brief Discussions on the Property of MS Model**

All what have been focused in this thesis is to make great efforts to achieve the best solving efficiency as possible as what we can. But still, people are not very aware of the reason of why this problem is so hard to be solved. This question may lead to the essentiality of simple investigation of the property of this model. So in this section, a little mathematical principle will be introduced briefly. With this knowledge, it will be helpful to bring some insights for understanding this problem deeper.

In integer programming theories, a technical term \( P \) is used to specify the problems that can be solved by *polynomial time algorithm*. *Polynomial time algorithm* can be understood in this simple way: the total running time for problem class \( P \) can be approximated by a polynomial function with regards to its instances. Mathematicians have proved linear programming problem belongs to \( P \). Interestingly; widely taught simplex method is not such *Polynomial time algorithm* but interior point method adopted by CPLEX is. As we know that GOM is a linear programming model, so no matter how big the size it is, it can always be expected to be done in “finite” time by CPLEX.

In addition, if a feasible problem \( X \) arises then \( P \) could change to \( NP \) problem. Feasible problem, simply saying, is a pair \((D, F)\) with \( F \subseteq D \), where the elements of \( D \) are finite binary strings. \( D \) is called the set of instances of \( X \), and \( F \) is called the set of feasible instances. Given an instance \( d \in D \), we want to determine whether \( d \in F \). The strict definition of \( NP \) is the class of feasibility problems such that for each instance with \( d \in F \), the answer \( d \in F \) is obtained in polynomial time by some nondeterministic algorithm. If in maintenance scheduling case, actually every instance from set \( D \) will be checked for feasibility, that is to say, those binary strings should not only meet the equation (3.32), also satisfy other constraints, such as elevation,
discharge or generation requirements. In this sense, MS problem has feasibility problem, so it
belongs to $NP$. Technically, there is one proposition being proved that “General integer
programming feasibility is in $NP$”.

In integer theories, the most difficult $NP$ problems are called the class of $d$ $NPC$ (NP-
Complete). $NPC$ is a subset of $NP$ problem, which is defined that if a problem belongs to $NP$ but
it could be solved by $P$. This causes a problem called transformability or more general,
polynomial reduction, which means that one problem can be solved in polynomial time given
that another one can. Without investigating too much into those theories, an easy understandable
conclusion is, the BIP feasibility problem belongs to $NPC$. Finally, a usually mentioned concept
called $NP$-hard is defined that if an $NPC$ problem has polynomial reducibility, then it is $NP$-hard.
MS problem is a mixed-binary programming model, which would be solved by being
transformed into any $P$ algorithm, such as CPLEX. Therefore MS belongs to $NP$-hard problem.
As Nemhauser (1988) pointed out in this book that, “If we know that the problem is $NP$-hard, we
can expect that some larger instances will be difficult for any algorithm.”

Above simple analysis could result in the conclusion that, MS is complicated enough to
be efficiently solved by any particular algorithm. The situation would be much worse if more
integer or binary requirements are added.