2010

Scenario reduction heuristics for a rolling stochastic programming simulation of bulk energy flows with uncertain fuel costs

Yan Wang
Iowa State University

Scenario reduction heuristics for a rolling stochastic programming simulation of bulk energy flows with uncertain fuel costs

by

Yan Wang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Industrial Engineering

Program of Study Committee:
Sarah M. Ryan, Major Professor
James D. McCalley
Sigurdur Olafsson
Leigh Tesfatsion
Lizhi Wang

Iowa State University

Ames, Iowa

2010

DEDICATION

I would like to dedicate this thesis to my son Benjamin Peiqi Yang, the constant inspiration of mine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Dr. Sarah M. Ryan has been the ideal thesis supervisor. Her sage advice, insightful criticisms, and patient encouragement aided the writing of this thesis in innumerable ways. All her advice and support for this project was greatly needed and deeply appreciated.

ABSTRACT

Stochastic programming is employed regularly to solve energy planning problems with uncertainties in costs, demands and other parameters. We formulated a stochastic program to quantify the impact of uncertain fuel costs in an aggregated U.S. bulk energy transportation network model. A rolling two-stage approach with discrete scenarios is implemented to mimic the decision process as realizations of the uncertain elements become known and forecasts of their values in future periods are updated. Compared to the expected value solution from the deterministic model, the recourse solution found from the stochastic model has higher total cost, lower natural gas consumption and less subregional power trade but a fuel mix that is closer to what actually occurred. The worth of solving the stochastic program lies in its capacity of better simulating the actual energy flows.

Strategies including decomposition, aggregation and scenario reduction are adopted for reducing computational burden of the large-scale program due to a huge number of scenarios. We devised two heuristic algorithms, aiming to improve the scenario reduction algorithms, which select a subset of scenarios from the original set in order to reduce the problem size. The accelerated forward selection (AFS) algorithm is a heuristic based on the existing forward selection (FS) method. AFS's selection of scenarios is very close to FS's selection, while AFS greatly outperforms FS in efficiency. We also proposed the TCFS method of forward selection within clusters of transferred scenarios. TCFS clusters scenarios into groups according to their distinct impact on the key first-stage decisions before selecting a representative scenario from each group. In contrast to the problem independent selection process of FS, by making use of the problem information, TCFS achieves excellent accuracy and at the same time greatly mitigates the huge computation burden.

CHAPTER 1 INTRODUCTION

1.1 Motivation

The modern lifestyle depends tremendously on the use and existence of fossil fuels, which are energy sources such as coal, oil and natural gas. Of the energy we use, 85% comes from fossil fuels. With this rate the world virtually depends on the supply of fossil fuel. But the common issue presented to us is that fossil fuels are running out. Owing to the increasing demand and limited availability of fossil fuels, the importance of efficient use of energy has been realized all over the world. Improving energy efficiency is a key strategy in making the world's energy system more economically and environmentally sustainable. The importance of energy efficiency lies in the fact that it ensures provision of same level of energy using smaller amounts of fossil fuels. Furthermore, reduced use of fossil fuels is essential in lowering the emission of greenhouse gases contributing to global warming.

Electrical power plants together with the production and transmission of fuels compose a complex large-scale network that involves many uncertain factors. Despite the inherent nonlinearities and uncertainties, remarkable efforts have been made to achieve a concise and comprehensive understanding of the large electric power network and to find more economic and reliable ways to organize and operate it. Due to the data availability and the complex interaction between subsystems, most energy models found in the literature have either a narrow geographic focus or a perspective limited to a single aspect of the whole system. Systems for the supply and transport of fuels and electric power therefore are investigated separately despite being highly interconnected. However, since 1974, the U.S. Department of Energy's Energy Information Administration (EIA) and its predecessor, the Federal Energy Administration (FEA), have developed a series of three computer-based, midterm energy modeling systems to analyze domestic energy-economy markets and the relationships among electric energy and all kinds of fuels.

EIA constructed huge energy models with sufficient data. However, from only publicly available sources, such as the websites of the EIA and the Canadian National Energy Board, Quelhas et al. [55] were able to formulate, validate and analyze a parsimonious and computationally efficient decision model to account for the medium term interdependencies across time and space in the U.S. bulk energy transportation system. This is a generalized minimum cost network flow model which is constituted by supply and storage of coal and natural gas (which together have accounted for approximately 70% of electricity generation in recent years [25]), electricity generation and the energy flows among these subsystems. The model can aid understanding of the tradeoffs between fuel transportation and electricity transmission as well as the ways in which fuel storage, fuel substitution and interregional electricity trade can be combined to meet the temporally and spatially variable demand for electricity, which cannot be stored in significant amounts. In a case study of the year 2002, the results indicated that the total cost of the fossil-fueled portion of the electricity system could be reduced considerably by relying far more heavily on generation from coal and increasing interregional trade [54]. However, the reliability of its conclusions could be limited by the lack of spatial and temporal detail necessitated by limitations in data availability. In addition, the deterministic model included an implicit assumption that all data for the year were known in advance. For fuel prices in particular, this assumption was inappropriate: the average price of natural gas was approximately 50% higher at the end of the year than its value predicted by EIA at the beigining of the year.

While the model offers related decision makers a comprehensive analysis of the national energy system, its formulation assumes that all information is known with certainty in advance. A question is raised by researchers: can we use this model to understand the effect of uncertainty on energy movements?

The reason to propose this question is quite natural given the uncertainty involved in the energy system. Multiple factors such as severe weather, equipment failures and international political events affect fuel prices, electric supply/demand and energy transportation. Some of the uncertain elements may cause a high cost to satisfy energy demands and some even lead to serious consequences; for example, large-scale disruption of energy supply. In 2005, hurricanes Katrina and Rita hit the Gulf of Mexico area. The catastrophic events not only interrupted the local electric and coal supplies but also damaged the natural gas production and transportation facilities, which caused significant nationwide impacts. The huge potential effects caused by the great uncertainty associated with the energy system motivate us to include uncertainty in the forecast elements within the model and study their effects using stochastic programming.

Stochastic programming has been applied to numerous energy models to address the problem of uncertain prices and demand. However, most of the research in the literature is limited to regional models or a single energy resource because of the spatial complexity and the interdependencies among various resources. Therefore, it would be interesting and meaningful to apply stochastic programming to the bulk energy transportation model and generate solutions to provide practical guidance for the U.S. power generation and transmission systems. A common issue in stochastic programming models is that they usually have a great number of scenarios which leads to large-scale optimization problems that can not be solved without techniques such as decomposition, sampling and scenario reduction. There is no doubt that we will encounter computational problems given the scale of this stochastic programming model. The complexity inspires our further investigation into computation relieving methods which either decompose the original problem into smaller sized problems and solve for the exact solution, or reduce the number of scenarios considered and aim at a good approximation. Although motivated by a specific application, these methodologies are suitable to a broader range of problems.

Finally, there is likely to be more uncertainty about supplies and greater environmental risk as less easily accessible reserves of fossil fuels are exploited. With levels of these fuels constantly decreasing, we should act now to become less dependent on fossil fuels and more dependent on renewable energy sources. Renewable energy is any natural source that can replenish itself naturally over a short amount of time. Renewable energy comes from many commonly known sources such as solar power and wind. A sustainable energy supply, both in the short- and the long-term, is needed for promoting both economic development and people's quality of life, as well as protecting the environment. The concepts of renewable energy and energy efficiency go hand in hand. To make the most of the sustainable energy policy there needs to be simultaneous application of strategies regarding renewable energy and efficient use of energy. These problems are generally large-scale problems with a lot of spacial and tempral complexities, which are similar to those incurred in the fossil fuel model. A example problem is making wind power storage decisions under random wind speed/direction. Hence, the stochastic programming methodology is appropriate for decision making with renewable energy by modeling the uncertain elements as discrete scenarios. And the lessons we learned from the fossil fuel model are also illuminating for generating and supplying renewable energy with high efficiency.

1.2 Objective

The goal of this dissertation is to first examine how the inclusion of uncertainty affects the model's results; in particular, in historical case studies, whether this inclusion improves the model's accuracy in simulation; and secondly, to develop a well validated heuristic scenario reduction algorithm that enhances computational efficiency for general large-scale stochastic programming models.

The energy system is fraught with uncertainty. The first question we need to answer is which factors to be modeled as stochastic elements and how to measure the uncertainty mathematically. We focus our study on uncertainty in fuel prices. Unlike natural disasters, they can be forecast with the aid of econometric models, but are less predictable than outages of generating units due to nonstationary influences. The difficulty of price forecasting is expected to increase as fossil fuels grow more scarce and regulations aimed at reducing carbon emissions are enacted more widely.

After the stochastic model is constructed, we should collect necessary data and draw solutions from the model with available methodologies that are used to address stochastic programming models. Since the solution in a historical case study is a simulation of energy movements, it will be compared to the actual flows for validation. Also, comparison between the stochastic technique and a deterministic approach weighs whether the additional computation cost for stochastic programming is worthwhile or not.

In order to find a solution to the large-scale optimization problem, we need to adopt well established methods such as decomposition and scenario reduction to reduce the problem size and the computation load. These methods are general and well accepted yet with restrictions such as the still heavy calculation burden and uncertain precision of the approximation. After applying and comparing the existing methods, it is worthwhile to study and develop algorithms that work particularly well in large-scale models with certain features possessed by our aggregated U.S. energy transportation network model with uncertain fuel prices.

In summary, the research objectives are to:

- Model the uncertain fuel prices as discrete scenarios and build a stochastic programming model for the bulk energy transportation system;
- Implement the model and construct case studies using historical data;
- Employ methodologies including aggregation and decomposition to solve the large-scale optimization optimization problem;
- Judge the value of the stochastic model by comparing the stochastic solution to both actual flows and the result from the deterministic model;
- State the lessons indicated by the stochastic model and its use in simulation of energy movements;
- Evaluate and comment on the accuracy and stability of the model using disaggregation and distribution perturbation.
- Apply a general scenario reduction algorithm and other methods that use the model features to find approximated solutions to the problem and compare the accuracy and efficiency of the alternatives.

- Propose and demonstrate a heuristic scenario reduction method that makes use of the problem specifications and provides tighter bounds on the discrepancy of the solution to the stochastic programming model than the general scenario reduction method.

1.3 Thesis organization

Chapter 2 reviews the relevant literature about integrated energy systems, stochastic programming in energy systems and computational issues. The formulation of the stochastic version of the energy transportation model is presented in Chapter 3 and a small numerical example illustrates the modeling methodology and some effects of fuel price uncertainty on the optimal decisions. Chapter 4 provides a detailed description of the model structure, data collection and the complete procedure for obtaining the solution of the optimization problem. Results of both the stochastic and deterministic models when tested with historical data for the years 2002 and 2006 are presented and compared in Chapter 5. Chapters 3–5 have been extracted from [71]. In Chapter 6 we employ three scenario reduction heuristics to increase computational efficiency [70]. Two heuristic scenario reduction methods with broader application are proposed and demonstrated in Chapter 7. Concluding remarks and directions for future work follow in Chapter 8.

CHAPTER 2 LITERATURE REVIEW

2.1 The national energy system

Due to the limited data availability and the complex interaction between subsystems, most energy models built in the literature are narrowed into contract/utility/region level and focus on one aspect of the whole system. Petroleum products, electric power, fuel supply and transmission systems are therefore investigated separately despite being highly interconnected. However, since 1974, the Federal Energy Administration (FEA) and its predecessor, the Energy Information Administration (EIA), have developed a series of three computer-based, medium term energy modeling systems to analyze domestic energy-economy markets and the relationship among electric energy and all kinds of fuels.

The Project Independence Evaluation System [33] was the first of the three systems and was employed by the FEA prior to 1982. It was initiated in 1974 to provide a framework for the developing of national energy policy through quantitative analysis and projections of the energy system. PIES considered several objectives including fuel price sensitivity, fuel competition (the possibility of the substitution of one energy source for another), technology restriction or improvement, resource limitations, economic impact, regional variations and other external effects on the energy system. Given the large volume of information and highly interdependent nature, a modular system was employed to permit the integration of subsystems, expansion of major components and the introduction of new elements. Figure 2.1 depicts the framework of PIES where the supply, demand, and equilibrium balancing components are combined with models of the economy, assessments of non-energy resource availability, and report writers that evaluate energy solutions in terms of the environmental, economic or resource impacts. PIES successfully analyzed the U.S. national energy system with an organization of engineering, econometric, and optimization models and improved the decision-making process for the complicated large-scale, time dependent system.

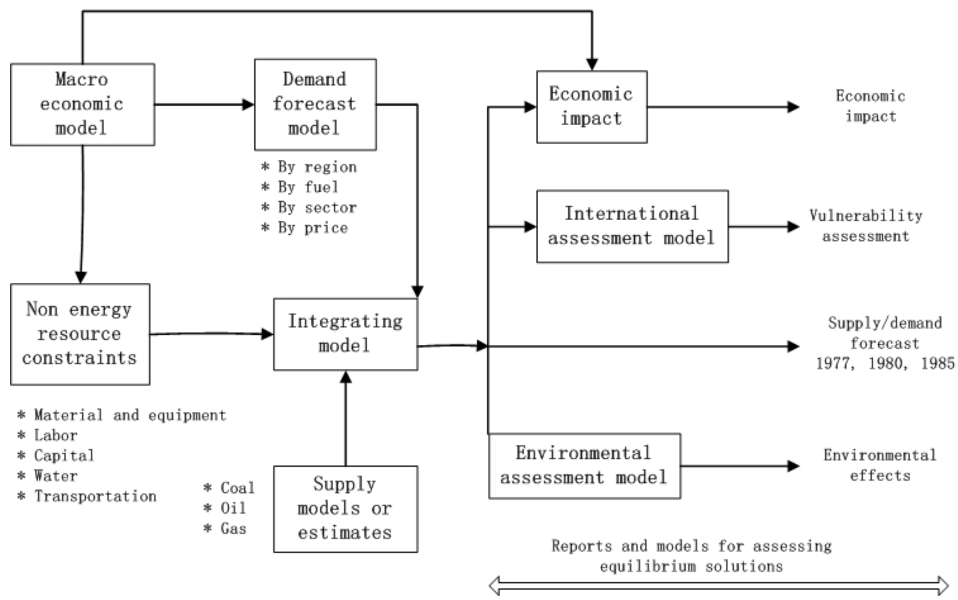

Figure 2.1 PIES structure [33]

In 1982, PIES was updated to the Intermediate Future Forecasting System (IFFS) which was used by EIA through 1993 [48]. While keeping the major objectives the same as the PIES, IFFS made a significant modification to the structure of model design, as shown in Figure 2.2. PIES built sub-models according to functions such as supply, demand and other constraints, keeping corresponding information about all the fuels in the same block. However, with the period of comprehensive energy legislation ending in the late 1970s, energy issues became more fuel specific, which motivated a model structured by fuels rather than functions. A simple integrating routine coordinates across the fuels and steps from submodel to submodel in order to capture the interaction among fuels. The new structure decomposes the model into manageable units which adopt diverse methodologies and are developed by individual groups with detailed knowledge of certain fuels. In contrast to the PIES in which the person responsible for the integrating methodology becomes unreasonably overburdened by the developmental runs needed to test changes in submodels, IFFS is partitioned by fuel to avoid the complex task of integration and to balance the workload among the staff in charge of submodels.

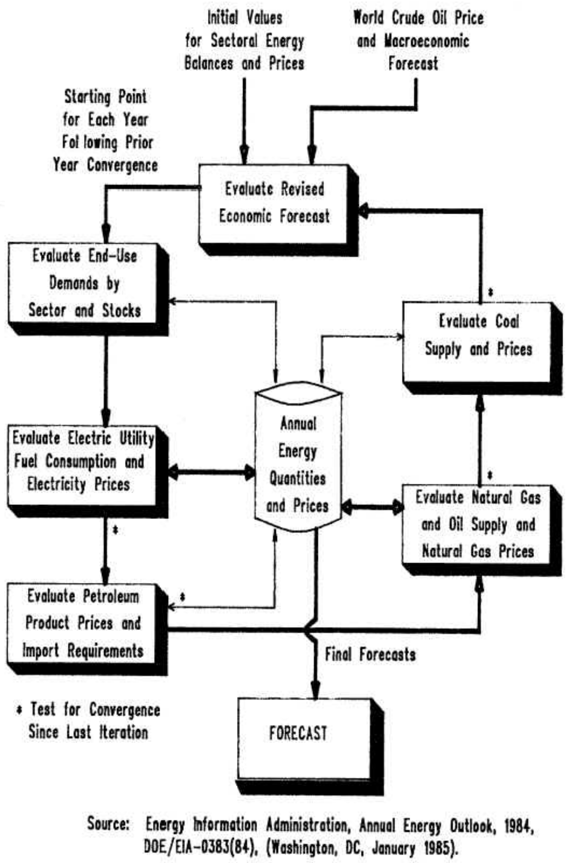

Figure 2.2 IFFS calculation flows (page 408 in [48])

In 1993, the IFFS was replaced by the National Energy Modeling System (NEMS) [23], which again had a new system structure. As depicted in Figure 2.3, it takes advantages from both PIES and IFFS. There are two levels of subsystems. The first level is composed by function components of Supply, Conversion and Demand. Within the function blocks of Supply and Conversion, submodels are built for individual fuels, while Demand is partitioned according to end-users. Associated with advanced modeling and optimization techniques and the latest computing machines, the NEMS combines and processes more energy information than its predecessors and therefore is more capable with projections. In addition to the baseline forecast published in the Annual Energy Outlook, NEMS generates one-time analytical reports and papers to analyze the effects of environmental impacts, existing government regulations and alternative energy policies. The system is used to test different assumptions about energy markets, as well as to evaluate the potential impacts of new and advanced energy production, conversion and consumption technologies. It has been used for special analysis at the request of the White House, U.S. Congress, other offices of the Department of Energy who specify the scenarios and assumptions, which means that the analysis produced by NEMS has an important effect on how the U.S. government regulates the energy markets. However, it is not open-source and not available for researchers and utilities to plan with.

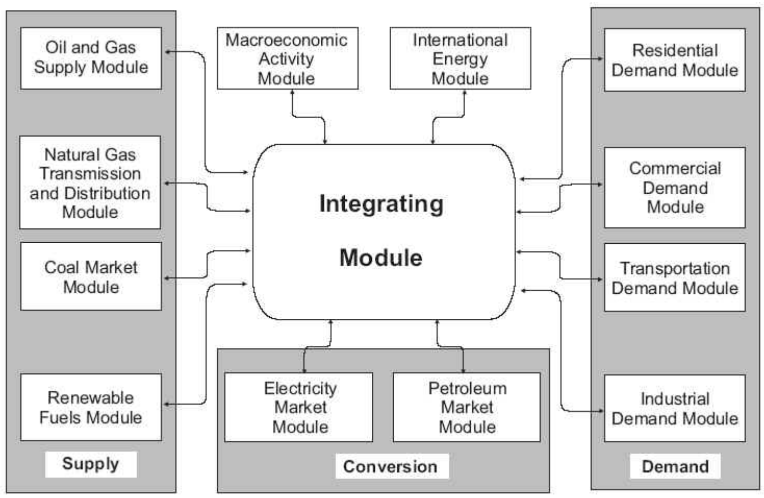

Figure 2.3 NEMS structure (page 10 in [23])

The Argonne National Laboratory developed a market-based simulation software program titled the Energy and Power Evaluation Program (ENPEP-BALANCE) [2]. Based on the input of energy system structure, energy production, consumption and prices, projected energy demand growth, and any technical and policy constraints, the nonlinear equilibrium model matches the demand for energy with available resources and technologies and determines the response of the energy system to changes in energy prices and demand levels (as shown in Figure 2.4). ENPEP-BALANCE's powerful graphical user interface makes it easy to build networks of regional, national, or multinational scope. It is free to anyone and is in use in over 80 countries. The model employs a market share algorithm to estimate the penetration of supply alternatives. The market share of a specific commodity is sensitive to the commodity's price relative to the price of alternative commodities. User-defined constraints (e.g., capacity limits), government policies (taxes, subsidies, priority for domestic resource over imported resource, etc.), consumer preferences, and the ability of markets to respond to price signals over time (i.e., due to lag times in capital stock turnover) also affect the market share of a commodity. It is reported [2] that the ENPEP-BALANCE approach simulates more accurately the complex market behavior of multiple decision makers that may not be captured by other optimization techniques that assume a single decision maker. Besides energy flows, the model's output includes environmental residuals such as greenhouse gases, waste generation, water pollution, and land use. The software is used extensively to conduct greenhouse gas mitigation analyses, energy policy studies, and natural gas market analyses.

Figure 2.4 The Energy and Power Evaluation Program [2]

Quelhas et al. [55] developed a generalized network flow model for the U.S. electric energy system to explore economic efficiency of the energy flows from fuel suppliers to electric load centers. Within this decision model, fuel production, transportation, storage, electricity generation and transmission are represented by nodes and arcs included in the generalized network which is a three-level system: Coal, natural gas and electricity are partitioned into corresponding levels and connected by energy movements among different levels. All the data in this model are derived from various public available sources, such as the websites of the Energy Information Administration and the Canadian National Energy Board. The model was validated by comparing its output to the actual data published by EIA for 2002 [54]. With the objective of cost minimization at the national level, the model is constrained by electricity generation/demand, fuel supply/demand and transmission capacities. It can be solved efficiently by network optimization codes and is expected to enable both public and private decision makers having limited available data and other resources to better understand the complex dynamics of interdependencies of primary fuels and electricity networks and carry out comprehensive analysis of a wide range of issues related to the energy sector.

2.2 Stochastic programming

Whereas deterministic optimization problems are formulated with known parameters, real world problems almost invariably include some unknown parameters. Randomness in problem data poses a serious challenge for solving many linear programming problems, through which the solutions obtained are optimal for the parameter estimates but may not be optimal for the situation that actually occurs. Stochastic programming (SP) is a framework for modeling optimization problems that involve uncertainty. This field is currently developing rapidly with contributions from many disciplines including operations research, mathematics, and probability. Conversely, it is being applied in a wide variety of subjects ranging from agriculture to financial planning and from industrial engineering to computer networks.

The fundamental idea behind stochastic programming is the concept of recourse, which was introduced by Dantzig [18] and Beale [3] independently. Recourse is the ability to take corrective action after a random event has taken place. The most widely applied and studied stochastic programming with recourse models are two-stage linear programs. Here the decision maker takes some action in the first stage, after which a random event occurs, affecting the outcome of the first-stage decision. A recourse decision can then be made in the second stage that compensates for any ill effects that might have been experienced as a result of the first-stage decision. The optimal policy from such a model is a single first-stage decision and a collection of recourse decisions defining which second-stage action should be taken in response to each random outcome. One natural generalization of the two-stage model is to extend it to many stages, each of which consists of a collection of recourse decisions corresponding to the set of possible realizations of the uncertain parameters up to that stage. Similarly to the single first-stage policy, the nonanticipativity requires that two realizations with identical values up to a certain stage must have identical decisions up to that stage. Mulvey and Vladimirou [45] specified stochastic programming to networks by dividing nodes and arcs into separate sets corresponding to the stage to which they belong. They also developed a scenario aggregation algorithm to maintain the network structure when decomposing a large-scale problem into small sub-problems.

An alternative type of stochastic programming approach is so-called chance-constrained programming, which was first introduced by Charnes and Cooper [15]. It does not require that decisions be feasible for every outcome of the random parameters. It seeks a decision which ensures that a set of constraints will hold with a certain probability. An application might be a delivery service that experiences random demands, and wishes to find the cheapest way to deliver its packages on time with a high probability.

While stochastic programming is usually characterized by a probability distribution on the parameters, robust optimization, which is a further development of chance-constrained SP, can tackle the problems where the parameters are only known within certain bounds. The goal is to find a solution that is feasible and acceptably close to optimal for all such data. Research with main contributions to the foundation of robust optimization includes Ben-Tal and Nemirovski [5] and Kouvelis and Yu [40]. Bertsimas and Sim [7] presented a robust optimization approach which set up a parameter to control the level of robustness against conservatism. This method provides a solution satisfying a high proportion (which depends on the parameter set) of the constraints even for the worst situation.

2.3 Scenarios in stochastic programming

A crucial step during the implementation of SP models is modeling the random parameters to well reflect the available knowledge on the randomness at hand. The uncertainties can be expressed in terms of multivariate continuous distributions or a discrete distribution with as many outcomes as necessary. In the formulation of a SP with discretely distributed parameters, the discrete scenarios are usually organized in the form of a scenario tree with nodes at levels which correspond to stages.

In the literature, there are generally three kinds of methods to generate a scenario tree. The first method is from data paths to the tree. Sufficient data paths are sampled from specified distributions or obtained from past observations. Mulvey and Vladimirou [43] developed a global scenario system which can simulate arbitrarily many sample paths using a calibrated model. The authors of [19], [36] and [62] use multivariate autoregression models to generate data paths in finance, hydroelectric power planning, and water resource applications, respectively. The next step is to build a scenario tree of a prescribed structure, which is determined by the horizon and stages of the model. This could be done by ad hoc crude methods, by cutting and pasting the data paths in a more or less intuitive way. Another option is the cluster analysis mentioned in [8] and [12]. It is also possible to generate the best approximating scenario tree of a given structure via a stochastic approximation technique suggested in [52].

The second choice of generating a scenario tree is directly generation of arcs sequentially. In [17], a Markov structure of data is exploited for conditional generation of scenarios in a way which takes into account the already created structure of the tree. A recent technique is sequential importance sampling by Dempster and Thompson [20]. In this algorithm, the tree is updated at every major iteration and addition sample paths are selected or previous realizations deleted, depending on the nodal importance estimates at the current iteration. The tree generated here is dynamic. The efficiency of the method depends on the adopted sampling rule.

The third algorithm, which was introduced in [34], is more flexible than the previous two in the sense that it can match more complex distributions with a limited number of constraints. One is free to specify any statistical properties he finds relevant. The method minimizes some measure of distance between the statistical properties of the generated outcomes and the specified values by solving a nonlinear program. While the method provides flexibility, the set of properties that are relevant is problem dependent and the choice of this set will affect the accuracy of model. Finally, there are also problem-oriented scenario trees. In [44] and [39], the trees are based solely on past observations. In other financial applications, such as [37] and [38], the scenario-based estimates of future asset prices are designed to not allow arbitrage opportunities.

In applications, the size of a scenario tree may grow extremely fast as the number of stages increases, producing an extraordinarily large SP problem which is difficult to solve due to memory limitations. Fortunately, we can find various effective methodologies in the literature to handle such large-scale linear programs. Benders decomposition [6] and other approaches derived from it are one series of schemes that decompose a large size problem into small subproblems. When Benders decomposition is applied to two-stage stochastic linear problems, the first stage is formulated as the master problem providing lower bounds, and a subproblem is formed for each scenario. All the subproblems together generate upper bounds and cuts for the master problem. The lower bound and upper bound eventually converge to the optimal solution. Benders decomposition keeps both the master and sub problems solvable. In effect, it substitutes computation time for memory requirements.

The drawback of decomposition is that it is time consuming to solve all the subproblems iteration by iteration given the large number of scenarios. Hence, sampling techniques are employed to reduce the number of sub problems. Lavenberg and Welch [41] discuss the efficiency of control variates in Monte Carlo sampling. Dantzig and Glynn [16] and Infanger [35] introduced importance sampling which is an improvement of Monte Carlo sampling. The approach embeds sampling in the decomposition algorithms where sampling only applies to the sub-problems so that there is no need to reformulate the whole problem according to each sample's results.

Compared to the repeating sampling procedure which might requires implementation of parallel computation, the scenario reduction method developed by Dupačová et al. [21] does a one-time reduction of the data paths or the scenario tree. Two heuristic algorithms are devised to select a subset of scenarios that has the minimum measure of distance to the original set of the scenarios. However, it is uncertain how accurately the selected scenarios would represent the original statistical specifications and how well the resulting solution approximates the true optimum. The general scenario reduction approach has come to be the standard built into the General Algebraic Modeling System (GAMS), which is a high-level modeling system for mathematical programming problems. There are also heuristics for certain types of problems. Carino et al. [13] choose scenarios according to desired mean and standard deviation. Beltratti et al. [4] separate the scenario tree into extreme scenarios and the most likely ones and a certain fraction of scenarios from each cluster are retained to represent the stochastic situation.

2.4 Stochastic programming models in energy

Stochastic programming models are widely used in the area of optimal allocation of energy and its related resources, where demand and prices are always unpredictable [68]. Those models in power systems planning are usually categorized according to the planning horizon. Long term planning models deal with a 15-20 year time horizon and consider large investments such as building thermal units and constructing hydro reservoirs and turbines. This kind of model helps us find the optimal investment to meet the uncertain future demand. Regularly, several possible future load duration curves are put forward and a straightforward recourse model is developed to address the recourse solutions for different scenarios. Murphy et al. [47] carried out a deterministic investment analysis using a new load duration curve carefully aggregated from scenario curves and obtained the same solution as if the recourse problem is solved. Sherali et al. [64, 65] emphasized peak load pricing and discussed Murphy's model in greater detail. Gardner and Rogers [30] investigated a multi-stage problem where load duration curves are revealed over time and investments are made stage by stage. While all the demand must be satisfied in traditional monopoly-based production planning, Qiu and Girgis [53] look at the problem from a different perspective by allowing and pricing outages, which takes into account that something even worse than the worst scenario modeled could occur with the consequence of shortage. Roh et al. [57] presents a stochastic program for generation and transmission expansion planning model in a competitive electricity market. Scenarios for random outages of generating units and transmission lines as well as long-term load forecasting are generated by Monte Carlo simulation. The scenario reduction technique of Dupačová et al. [21] is introduced for reducing the computational burden of a large number of planning scenarios.

Medium-term power planning has a 1-3 year horizon. Many applications deal with hydro reservoir management, where the true cost and risk brought by the uncertain aspects of using the water are underestimated by deterministic solutions. The performance of stochastic optimization models is proved to be significant. Pereira and Pinto [51] applied a multi-stage stochastic dydamic programming methodology to a system of 39 hydroelectric plants in order to solve for the optimal generation plan, which is composed of 10 monthly decisions. In this model the future inflows sequences, of which it is impossible to get a perfect forcast, are the stochastic parameters and the objective is to minimize the expected-cost-to-go functions which are piecewise linear. Butler and Dyer [11] used a linear two-stage stochastic programming model to aid a utility company in making optimal natural gas purchase decisions under uncertain future gas prices and demands. Within the one year horizon considered, in order to reduce the scale of problem, the later periods are aggregated and a daily-weekly-monthly model was adopted. Ryan and Wang [61] reviewed stochastic models in energy transportation and storage for dispatchable power generation. Fuel and electricity prices and future energy demands are the most common uncertainties that need to be considered.

All of the above models were developed for regulated markets. The transition of electricity markets from the old regulated regime to the deregulated system motivated the development of hybrid stochastic models where there is both a demand constraint and a wholesale market, where the producer can choose to serve the local load by his own production capacity or by buying capacity. Some stochastic programming models serve the needs of utility planners and policy makers in that they can generate scenarios for market prices of electricity. Important papers include [29], [10] and [32].

As the electricity markets are developing into regional commodity markets, the use of standardized financial contracts such as forward contracts increases. The contract price represents the current market value of future delivery of the electricity. Hence, valuation of future production is needed in stochastic programming models in energy. These models are based on describing the uncertainty in the form of scenarios of the spot price of the commodity. Although basing the scenarios on forecasts of spot prices will not give a valuation that is consistent with the market, the stochastic programming models can value the decision flexibility using a price of risk that is consistent with the market.

Short term planning typically deals with problems having horizons of one week or shorter, such as unit commitment (UC) and economic dispatch. Wang et al. [69] built a SP model to solve a single day security-constrained UC problem with uncertain wind power generation. The scenario reduction technique is employed to reduce the large number of sampled scenarios. The strategy of rolling planning with scenario trees is adopted in [66] to solve a similar UC problem with a time horizon of 36 hours. Scenario reduction is also applied. Energy bidding is viewed as a short term optimization problem in which the market participant offers to buy or sell capacity to the market in the form of price-quantity pairs for given time intervals. Determining optimal bids to send to the market operator becomes a nontrivial task that can be supported by stochastic programming models. Nowak and Römisch [50] studied this problem and presented an integrated stochastic unit commitment and bidding model. Neame et al. [49] and Anderson and Philpott [1] also developed stochastic models to explore optimal energy bidding prices.

Operations scheduling in deregulated markets is divided into two categories. In the first set of problems, generation utilities are not large enough to influence electricity prices by changing the amount of generation capacity offered to the market. Scott and Read [63] investigated the other class of models in which the operators do have market power on energy price. A major limitation in these analyses is that buying and selling of contracts is in reality determined simultaneously with production.

Financial instruments such as forward contracts are used to reduce risk in energy markets. However, because fixing income in the future does not automatically mean reduced risks, researchers made great efforts on stochastic models that manage the risk of energy trading. Mo et al. [42] and Fleten et al. [28] suggest that production scheduling and contract risk management should be integrated in order to maximize expected profit at some acceptable level of risk. However, other researchers claim that the benefits of a decoupled set of models will probably outweigh the small theoretical gain from integrating production planning and trading. All in all, the deregulated markets have not found their final forms and there are many more topics that can be investigated with the tool of stochastic programming.

2.5 The gap in the literature

In this chapter, we reviewed the existing national energy systems, stochastic programmming and its applications in energy systems. There is no doubt that a range of methodologies are involved in the complex national energy models constructed by EIA and others to address the uncertainties in the huge system. However, details of these models are not revealed. The modeling of uncertainty in medium-term national energy systems is still largely missing in the literature. In this report, we will include uncertain fuel prices in Quelhas's model [55], use the stochastic programming method to model the uncertainty and better simulate the energy flows in national scope. In addition to the complexity of subsystems which is the main reason for lack of national wide energy models in the literature, the tremendous computation task encountered in the large-scale optimization problem also matters. The size of a model usually grows dramatically when uncertainty is considered. Therefore, we will also address the computational efficiency issue that is inherent with the large-scale stochastic model.

CHAPTER 3 A STOCHASTIC PROGRAM FOR FUEL COST UNCERTAINTY

3.1 Deterministic model and its notations

Our model of the U.S. national electric energy system is aggregated by regions based on the high-level topology of the electrical interconnections and fuel transportation infrastructure. It is an adequate simplification of the physical and institutional complexity of the electric power industry given that data are generally available at this level [56]. The whole system is modeled as a generalized minimum cost flow network. The nodes represent entities such as coal mines, natural gas wells, natural gas storage facilities and electricity demand centers at different time periods. The flows among these nodes include the transportation or storage of fuel and the transmission or regional trade of electricity. The flow multipliers quantify transmission or transportation losses and the efficiency of conversion from fuel to electric energy. The mathematical formulation of this model is as equation (3.1) [55]. Table 3.1 shows notations used in the formula.

$$
\begin{array}{lll}
\min_e & \sum_{(i,j;t)\in A} c_{ij}(t)e_{ij}(t) & \\
\text{s.t.} & \sum_{(j,k;t)\in A} e_{jk}(t) - \sum_{(i,j;t)\in A} r_{ij}(t)e_{ij}(t) = b_j(\tilde{t}) & \forall j \in N, \forall \tilde{t} \in T \\
 & l_{ij} \leq e_{ij}(t) \leq u_{ij} & \forall (i,j;t) \in A
\end{array}
\tag{3.1}
$$

3.2 Stochastic model

We investigate the impacts of uncertain fuel cost. It is reasonable to formulate these quantities as discrete random variables taking a finite number of realizations. The assumption of discrete distributions for the uncertain elements is common in most stochastic programming

Table 3.1 Notations in deterministic model

A	Set of arcs.
N	Set of nodes.
T	Set of time periods.
t	Index for time periods.
$(i, j; t)$	The arc from node i to node j in period t.
$e_{ij}(t)$	The decision variables, which represent energy in the form of fuels or electricity flowing from node i to node j during period t.
$b_j(t)$	Supply (if positive) or demand (if negative) at node j during time t.
u_{ij}	Upper bound on the energy flowing from node i to node j.
l_{ij}	Lower bound on the energy flowing from node i to node j.
$c_{ij}(t)$	Per unit cost of the energy flowing from node i to node j during time t.
$r_{ij}(t)$	Flow multiplier associated with the arc from node i to j during time t.

models. We model the cost per unit flow on a fuel acquisition arc as a random variable such that $Pr\{c_{ij}(t) = c_{ij}^k(t)\} = p_{ij}^k(t), k = 1, ..., K$. Given a total of m random cost variables over the problem horizon, we can define a scenario $s \in S$ as an m-vector of values that occur jointly with probability π_s.

When applied to a generalized network flow problem, the two-stage approach requires that all the arcs and nodes be divided into two sets [45]. The set of arcs A_1, on which the flows have to be decided before the uncertain quantities are revealed, are the first-stage arcs and the set of arcs A_2, on which decisions are made after, are included at the second-stage. In our model, if the current period is $\hat{t}$, then $A_1(\hat{t}) = \{(i, j; t) \in A, t = \hat{t}\}$ and $A_2(\hat{t}) = \{(i, j; t) \in A, t > \hat{t}, t \in T\}$; Let $\Delta_i^+ = \{(i, j; t) \in A\}$ and $\Delta_i^- = \{(j, i; t) \in A\}$. The nodes are partitioned into sets: $N_1(\hat{t}) = \{i : \Delta_i^- \cup \Delta_i^+ \in A_1(\hat{t})\}$ and $N_2(\hat{t}) = N \backslash N_1(\hat{t})$.

The notation $x_{ij}(t) = e_{ij}(t), (i, j; t) \in A_1(\hat{t})$, and $y_{ij}(t) = e_{ij}(t), (i, j; t) \in A_2(\hat{t})$, distinguishes between first stage flows and second stage flows. All the scenarios are considered jointly in the solution procedure. Because the values of the first-stage decisions must be invariant over all scenarios, we have $z_{ij}(t) = x_{ij}^s(t), \forall s \in S(t), \forall (i, j; t) \in A, t = \hat{t}$. Therefore, the overall problem to minimize expected cost at period $\hat{t}$ can be stated as the deterministic equivalent (3.2), where $\hat{t}$ is suppressed in the notation of A_1, A_2, N_1 and N_2.

$$
\begin{aligned}
\min_{(z,y)} \quad & \textstyle\sum_{(i,j;t)\in A_1} c_{ij}(t) z_{ij}(t) + \sum_{s\in S(\hat{t})} \pi_s \sum_{(i,j;t)\in A_2} c^s_{ij}(t) y^s_{ij}(t) \\
\text{s.t.} \quad & \textstyle\sum_{(i,j;t)\in \Delta_i^+} z_{ij}(t) - \sum_{(j,i;t)\in \Delta_i^-} r_{ji}(t) z_{ji}(t) = b_i(\tilde{t}) \qquad \forall i \in N_1 \\
& \textstyle\sum_{(i,j;t)\in \{\Delta_i^+ \bigcap A_1\}} z_{ij}(t) - \sum_{(j,i;t)\in \{\Delta_i^- \bigcap A_1\}} r_{ji}(t) z_{ji}(t) + \\
& \textstyle\sum_{(i,j;t)\in \{\Delta_i^+ \bigcap A_2\}} y^s_{ij}(t) - \sum_{(j,i;t)\in \{\Delta_i^- \bigcap A_2\}} r_{ji}(t) y^s_{ji}(t) = b_i(\tilde{t}) \quad \forall i \in N_2, \forall s \in S(\hat{t}) \\
& l_{ij}(t) \le z_{ij}(t) \le u_{ij}(t) \quad \forall (i,j;t) \in A_1 \\
& l_{ij}(t) \le y^s_{ij}(t) \le u_{ij}(t) \quad \forall (i,j;t) \in A_2, \forall s \in S(\hat{t})
\end{aligned}
\tag{3.2}
$$

Say $|A_1| = n_1$ and $|A_2| = n_2$, $|N_1| = m_1$ and $|N_2| = m_2$. The total number of different scenarios is $|S(\hat{t})|$. Hence, the size of this deterministic equivalent formulation is $m_1 + |S(\hat{t})| m_2$ arcs and $n_1 + |S(\hat{t})| n_2$ nodes.

The solution to a two-stage stochastic program such as (3.2) is called the *recourse problem (RP) solution*. To analyze how the decisions are affected by including uncertainty and using stochastic programming (SP), we can compare the results of the stochastic model to three alternatives. A common approach to decision-making in an uncertain environment is to substitute each random variable by its expected value and solve the resulting deterministic problem (3.1). The *expected value (EV) solution* is the result where the first-stage variables have been fixed at their values obtained by solving this deterministic problem and the second-stage variables vary according to each scenario. Both *RP* and *EV* are usually compared to *wait and see (WS) solution*, which is the collection of the optimal solution to each scenario. Besides *EV* and *WS*, we use the *true value (TV) solution*, which is defined as a single set of optimal (x, y) given the true values of the uncertain parameters, to match the real case study in chapters 4 and 5 where the actual costs of natural gas do not fall into any of the forecasted scenarios.

3.3 Small example and solutions

Before introducing the rolling horizon simulation, to understand possible effects of uncertain fuel cost, we apply the two-stage approach to a two-period illustration of an integrated energy network with one coal-fired plant, one natural gas (NG) plant and one electricity demand center. Through this example, we illustrate how fuel storage and diversity of fuel supply increase when uncertainty is considered. Moreover, although RP is closer to WS than EV is, it could differ from TV more than EV does, depending on how closely the true values correspond to the forecast scenarios.

The network is shown in Figure 3.1, where for simplicity all flows are in equivalent units of electricity. The fuel suppliers (coal mine and gas well) are integrated with the power generators so that the costs associated with the arcs from the plants to the demand center include the fuel cost, generation cost and transmission cost. Note that c_s represents the cost of both acquiring and storing natural gas. Nodes "1" and "2" represent the demand center in the first and second periods, respectively.

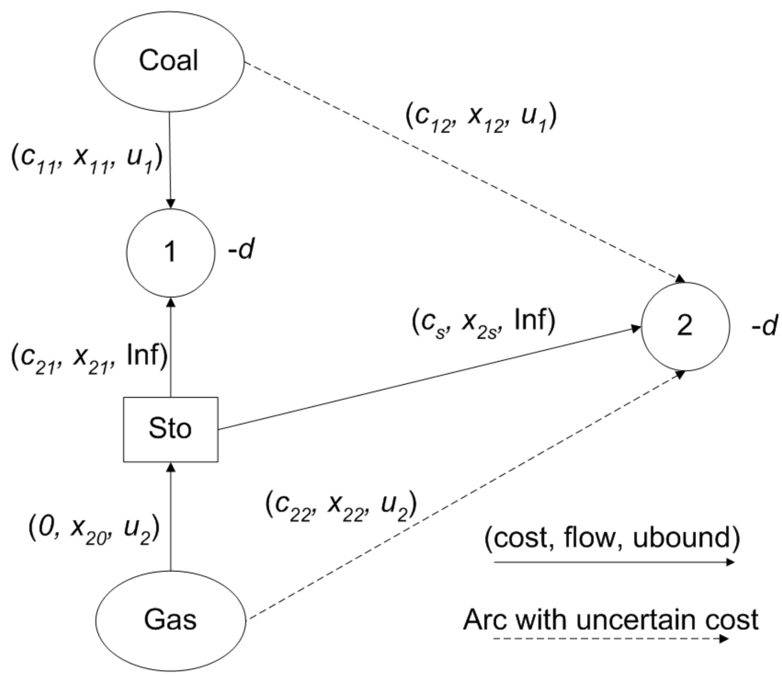

Figure 3.1 A two-period small example

We consider a stochastic problem where the costs of fuels in the second period are uncertain. There are two possible values for each random variable, which are considered independent. The four scenarios are listed in Table 3.2.

Assume $u_2 < u_1 < d < u_1+u_2$, $c_{11} = c_{12}^L < c_{22}^L < c_{12}^H < c_{22}^H$, and $c_{11} < c_{21} < c_s < c_{22}^H$. Given

Table 3.2 Scenarios of the two-period small example

Scenario	Coal Cost	NG Cost	Probability
s^1(LL)	$c_{12}^1 = c_{12}^L$	$c_{22}^1 = c_{22}^L$	$\pi_1 = p_1 p_2$
s^2(HL)	$c_{12}^2 = c_{12}^H$	$c_{22}^2 = c_{22}^L$	$\pi_2 = q_1 p_2$
s^3(LH)	$c_{12}^3 = c_{12}^L$	$c_{22}^3 = c_{22}^H$	$\pi_3 = p_1 q_2$
s^4(HH)	$c_{12}^4 = c_{12}^H$	$c_{22}^4 = c_{22}^H$	$\pi_4 = q_1 q_2$

Table 3.3 *RP* and *EV* solutions for the small example

	RP				*EV*			
	s^1(LL)	s^2(HL)	s^3(LH)	s^4(HH)	s^1(LL)	s^2(HL)	s^3(LH)	s^4(HH)
x_{11}	u_1	u_1	u_1	u_1	u_1	u_1	u_1	u_1
x_{20}	u_2	u_2	u_2	u_2	$d-u_1$	$d-u_1$	$d-u_1$	$d-u_1$
x_{21}	$d-u_1$	$d-u_1$	$d-u_1$	$d-u_1$	$d-u_1$	$d-u_1$	$d-u_1$	$d-u_1$
x_{2s}	u_1+u_2-d	u_1+u_2-d	u_1+u_2-d	u_1+u_2-d	0	0	0	0
x_{12}	u_1	$2d-u_1-2u_2$	u_1	u_1	u_1	$d-u_2$	u_1	u_1
x_{22}	$2d-2u_1-u_2$	u_2	$2d-2u_1-u_2$	$2d-2u_1-u_2$	$d-u_1$	u_2	$d-u_1$	$d-u_1$
Coal	$2(1-\pi_2)u_1 - 2\pi_2 u_2 + 2\pi_2 d$				$(2-\pi_2)u_1 - \pi_2 u_2 + \pi_2 d$			
Gas	$2\pi_2 u_2 - 2(1-\pi_2)u_1 + 2(1-\pi_2)d$				$\pi_2 u_2 - (2-\pi_2)u_1 + (2-\pi_2)d$			

the expected cost of fuel f in period 2, $\bar{c}_{f2} = p_f c_{f2}^L + q_f c_{f2}^H$, $f = 1, 2$, where $q_f = 1 - p_f$, the *EV* solution (Table 3.3) is obviously to use as much coal as possible because $c_{11} < \bar{c}_{12} < \bar{c}_{22} < c_s$.

The uncertain costs will be revealed at the beginning of the second period. An arc is included in the first stage if the flow on that arc is decided before c_{12} and c_{22} are known. Here $A_1 = \{(Coal, 1), (Gas, Sto), (Sto, 1), (Sto, 2)\}$ and $A_2 = \{(Coal, 2),$
$(Gas, 2)\}$.

The stochastic program can be formulated as (3.3) and the complementary slackness conditions are listed in (3.4).

$$\min_x \quad c_{11}x_{11} + c_{21}x_{21} + c_s x_{2s} + \pi_1(c_{12}^1 x_{12}^1 + c_{22}^1 x_{22}^1) + \pi_2(c_{12}^2 x_{12}^2 + c_{22}^2 x_{22}^2)$$

$$+\pi_3(c_{12}^3 x_{12}^3 + c_{22}^3 x_{22}^3) + \pi_4(c_{12}^4 x_{12}^4 + c_{22}^4 x_{22}^4)$$

$$\text{subject to} \qquad -x_{11} \geq -u_1 \qquad [v_{11}]$$

$$-x_{12}^i \geq -u_1 \qquad [v_{12}^i, i = 1, 2, 3, 4]$$

$$
\begin{aligned}
-x^i_{20} &\geq -u_2 && [v_{20}] \\
-x^i_{22} &\geq -u_2 && [v^i_{22}, i = 1,2,3,4] \\
x_{11} + x_{21} &= d && [w_1] \\
x_{20} - x_{21} - x_{2s} &= 0 && [w_0] \\
x^i_{12} + x_{2s} + x^i_{22} &= d && [w^i_2, i = 1,2,3,4] \\
x_{11}, x^i_{12}, x^i_{22}, x_{20}, x_{21}, x_{2s} &\geq 0, && i = 1,2,3,4
\end{aligned}
\tag{3.3}
$$

$$
\begin{aligned}
x_{11}(c_{11} + v_{11} - w_1) &= 0 \\
x_{20}(v_{20} - w_0) &= 0 \\
x_{21}(c_{21} - w_1 + w_0) &= 0 \\
x^i_{12}(\pi_i c^i_{12} + v^i_{12} - w^i_2) &= 0, \quad i = 1,2,3,4 \\
x_{2s}(c_s - w^1_2 - w^2_2 - w^3_2 - w^4_2 + w_0) &= 0 \\
x^i_{22}(\pi_i c^i_{22} + v^i_{22} - w^i_2) &= 0, \quad i = 1,2,3,4 \\
v_{11}(u_1 - x_{11}) &= 0 \\
v_{20}(u_2 - x_{20}) &= 0 \\
v^i_{12}(u_1 - x^i_{12}) &= 0, \quad i = 1,2,3,4 \\
v^i_{22}(u_2 - x^i_{22}) &= 0, \quad i = 1,2,3,4 \\
v_{11}, v^i_{12}, v_{20}, v^i_{22} &\geq, \quad i = 1,2,3,4
\end{aligned}
\tag{3.4}
$$

Considering the first stage variables, we find $x_{11} = u_1$,$x_{21} = d - u_1$ because $c_{11} < c_{21}$, regardless of the period 2 costs. But the optimal values of x_{20} and x_{2s} are not clear. When the fuel costs in the second period are high, it is beneficial to use storage as much as possible. However, it is also possible that the NG cost will be low and storing gas becomes relatively expensive.

Consider the possibility of having $x_{20} = u_2$ in the optimal *RP* solution. According to the assumptions of fuel costs and scenarios, $x_{12}^1 = x_{12}^3 = x_{12}^4 = u_1$, $x_{22}^1 = x_{22}^3 = x_{22}^4 = 2d - 2u_1 - u_2$, $x_{12}^2 = 2d - u_1 - 2u_2$ and $x_{22}^2 = u_2$. Therefore, $w_2^i = \pi_i c_{22}^i$ for $i = 1, 3, 4$ and $w_2^2 = \pi_2 c_{12}^2$. From the complementary slackness condition $c_s - w_2^1 - w_2^2 - w_2^3 - w_2^4 + w_0$, for $x_{20} = u_2$ to be optimal it is necessary that $\bar{c}_{22} - c_s + \pi_2(c_{12}^H - c_{22}^L) \geq 0$. It can be shown that the strict version of this inequality is sufficient for $x_{20} = u_2$ to be optimal.

Table 3.3 compares the *RP* and *EV* solutions. Taking the expectation over scenarios, *RP* uses less coal and more natural gas than *EV* because it stores natural gas while *EV* does not. In particular, it is able to exploit the low price of gas relative to coal in scenario 2. The results from the stochastic program promote more NG storage and reduce advance commitments to coal, indicating that the introduction of uncertain fuel costs leads to the diversification of fuel supply. Moreover, the *RP* solution holds under the condition of $\bar{c}_{22} - c_s + q_1 p_2(c_{12}^H - c_{22}^L) > 0$ regardless of the exact specification of p_1 and p_2; therefore, within a certain range, the stochastic solution is not sensitive to the distributions of the random parameters.

Consistent with the traditional perception, *RP* is closer to *WS* (Table 3.4) than *EV* is, in that *RP* and *WS* use the same expected amount of each fuel (the only difference is storage vs. purchase in the 2nd period). However, the comparison may be different for *TV* and *RP*. As mentioned before, the actual costs can deviate from all of the scenarios. Here, two cases are considered. In Case E, the actual costs are consistent with the expectation that $c_{12} < c_{22} < c_s$. Case H describes a very extreme situation where the cost of coal is very low and the cost of NG is very high, such that $c_{12} < c_s < c_{22}$. Table 3.5 presents the *TV* solutions for both cases. The *RP* solution could differ from *TV* in terms of the first stage variables more than *EV* does when the true values are close to the expectations as in Case E. But when the costs deviate from the expected values as in case H, the first stage decisions of *RP* are more similar to those of *TV* because the stochastic method makes use of the information from those extreme scenarios. The same qualitative comparisons, though less extreme, will be seen in results for 2002 and 2006 in Chapter 5.

Table 3.4 *WS* solution for the small example

	s^1(LL)	s^2(HL)	s^3(LH)	s^4(HH)
x_{11}	u_1	u_1	u_1	u_1
x_{20}	$d-u_1$	$d-u_1$	u_2	u_2
x_{21}	$d-u_1$	$d-u_1$	$d-u_1$	$d-u_1$
x_{2s}	0	0	u_1+u_2-d	u_1+u_2-d
x_{12}	u_1	$d-u_2$	u_1	u_1
x_{22}	$d-u_1$	u_2	$2d-2u_1-u_2$	$2d-2u_1-u_2$
Coal	$2(1-r_2)u_1-2r_2u_2+2r_2d$			
Gas	$2r_2u_2-2(1-r_2)u_1+2(1-r_2)d$			

Table 3.5 *TV* solutions for Case H and Case E

	TV_E	TV_H
x_{11}	u_1	u_1
x_{20}	$d-u_1$	u_2
x_{21}	$d-u_1$	$d-u_1$
x_{2s}	0	u_1+u_2-d
x_{12}	u_1	u_1
x_{22}	$d-u_1$	$2d-2u_1-u_2$
Total Coal	$2u_1$	$2u_1$
Total NG	$2(d-u_1)$	$2(d-u_1)$

3.4 The rolling procedure

While the two-stage approach is natural for two-period problems, the model introduced in section 3.1 has a horizon of one year with 12 periods. If the decision model were to be used for planning, we could adopt a multi-stage counterpart of the two-stage approach and formulate the evolution of uncertain fuel costs in terms of a standard multi-stage scenario tree. Such a formulation would allow the planning decisions in each stage to depend on realizations of uncertain quantities as they unfolded over time according to the fixed scenario tree defined at the outset. However, here a rolling two-stage approach is employed instead. Rather than future planning, we use the model here for historical simulation; hence, the uncertain fuel costs are modeled according to how the forecasts actually evolved over time. The actual prices in a period did not coincide with any of the scenarios defined by forecasts from the previous period, and the forecasts of future periods were revised each period. Thus, for example, the price realized at period 3 differed from all scenarios defined in period 2, and the scenarios defined in period 3 for period 4 also differed from those defined in period 2 for period 4. This evolutionary description of uncertainty motivated the use of a rolling horizon simulation. To simulate the actual decision making process with forecast updates, the stochastic program was reformulated and solved repeatedly, each time solving for the current/first period decisions with a collection of newly updated scenarios for the remaining periods.

When solving a problem with periods from t_0 to T, we start with $\hat{t} = t_0$ and obtain a solution (z, y) from(3.2), among which only first-stage decisions z for the current period t_0 are kept and the elements of z are removed from the set of decision variables. At the beginning of period t_0+1, set $\hat{t} = t_0 + 1$ and the two sets of arcs are $A_1(t_0 + 1)$ and $A_2(t_0 + 1)$. The true values of $c_{ij}(t_0+1)$ are revealed and $S(t_0+1)$ contains the scenarios for $\{c_{ij}(t), t \geq t_0+1\}$ in line with the new forecast. The decisions for period t_0+1 are obtained by solving the renewed problem (3.2). Consequently, the full recourse problem solution $\{e_{ij}(t), t \in T\}$ is completed as we simulate decision-making in all periods. In accordance to the rolling decision making procedure which retains only the first-stage decision of each recourse problem solution, the RP solution here is the collection of first-stage decisions $\{z(1), z(2), ..., z(T)\}$. Similarly to RP but instead of the complex SP formulation (3.2), the expected value (EV) solution $\{\bar{x}(1), \bar{x}(2), ..., \bar{x}(T)\}$ is the collection of first period decisions for a series of deterministic problems (3.1) with actual prices replaced by the mean value of the forecasts. Note that the forecasted average costs are also updated each period, thus rolling is necessary even for the deterministic formulation. A simple three-period example in appendix A illustrates the procedures of obtaining RP, EV and TV.

Whereas we could apply the rolling procedure on a multi-stage formulation, we chose the two-stage formulation, which is a relaxation of the latter by removal of non-anticipativity constraints on decisions after the first period [59]. Such constraints are not crucial in our problem because the later period decisions are discarded and only the first period decisions for each horizon are retained upon each roll forward. This kind of relaxation has been successfully adopted in a real world energy planning problem [11] where the utility focuses on the current/first period contract and operating decisions. Another important advantage of the two-stage approach is its ease of formulation and solution by decomposition. A multi-stage model would require much more effort to reconstruct the tree upon each roll forward, and to decompose the problems with special techniques. It is unclear whether the extra work is worthwhile due to the nature of forecast revision and uncertainty resolution in the simulation.

CHAPTER 4 IMPLEMENTATION

4.1 Model validation

In the network model of the U.S. bulk energy system, actual coal mines and natural gas wells are aggregated regionally into 11 coal supply nodes and 14 natural gas supply nodes based on 2002 data. See the appendix for the lists of nodes (Table B.1) and arcs (Table B.2), and the network graphs (Figure B.1 to Figure B.4) for each subsystem and the integrated system. Every node is characterized by its productive capacity and average minemouth/wellhead price. For coal mines, the average heat value and average sulfur content are also included. The 17 nodes representing electric demand centers correspond to the demand regions defined by North American Electric Reliability Corporation (NERC) among which electricity is traded and transferred (reorganized to a set of 15 by 2006). For each demand region, energy generation plants are aggregated to a single node if they use the same fuel type and prime mover. There are 6 different types of plants and a total of 102 generation nodes in the system. Each generation node is assigned combined capacity and weighted average heat rate for all the power plants of a certain type in that region. Arcs are established between fuel supply nodes and the matching generation nodes. These arcs are characterized by transportation capacity, cost and efficiency. Gas wells are connected to storage facilities and withdrawal capacities and the storage costs are assigned to the corresponding arcs. While the gas can be carried over from one period to the next, nodes denoting the same storage facility in consecutive periods are connected by an arc with a lower bound representing the cushion gas and an upper capacity bound. Within each demand region, power flows from the 6 power plants to the demand center. There is also electricity transmitted among demand centers. All the power transmission paths have their own capacity bounds, costs and loss factors. With year 2002 data, using monthly natural

Table 4.1 Total flows comparison: 2002 actual data and the model

Result	Actual	Case A	Case B
Coal deliveries (million tons)	976	953	1,054
NG deliveries (million Mcf)	5,398	5,125	3,615
Net electric power trade (thousand GWh)	N/A	205	382

gas and electricity nodes and yearly coal nodes, there are totally 1290 nodes, 3480 arcs in this deterministic model. Note that the model includes energy balance but not power flow constraints. Ryan et al. [60] include the latter constraints as well as strategic generator behavior and a market-balancing system operator with a highly simplified fuel network.

Quelhas et al. [54] verified the model (3.1) by comparing results from the model to actual aggregated flows. As shown in Table 4.1, the first column contains actual coal and NG deliveries in year 2002 and the other two columns contain total flows calculated from the model. In Case A, optimized coal and NG flows are solved by fixing generation and demand to the actual data at each electricity demand center, while Case B is solved with only demands fixed. The small difference between Case A and the actual data validates the model of the fuel subsystems and conversion to electric energy. Comparing Case A to Case B, the optimal flows indicate that greater economic efficiency could be achieved if more electricity were generated from coal and more electricity were traded among subregions.

4.2 The two stage decisions and scenario generation

The long term fuel cost graph in Figure 4.1 is taken from page 64 in 2006 EIA Annual Energy Review [25]. Because the coal price is quite flat, it is treated as fixed. The natural gas (NG) price is much more variable and therefore treated as an uncertain cost in the stochastic model. Given the distinct levels of price stability, generators usually make long term coal contracts and short term NG contracts. In our model, it is assumed that we set up a single coal contract at the beginning of the year while NG purchases and power generation decisions are made at the beginning of each month. Therefore, time step for the coal subsystem is one year while it is one month for the NG and electrical subsystems. Unlike the usual rolling horizon methods where the length of simulation interval remains constant, the distinct time steps for individual subsystems makes a constant simulation horizon impractical. Instead, the length of each succesive horizon is reduced by one period. In January, the first-stage decisions include the coal purchase of the whole year, NG purchase and electricity generation for the first month. The uncertainties consist of all the future NG costs, which affect the current decision through fuel storage. After January decisions are made, the second month NG price is known, NG purchase and electricity generation for February become the first-stage decisions, leaving rest of the decisions in the second-stage. The problem is completely solved when December's decisions are obtained.

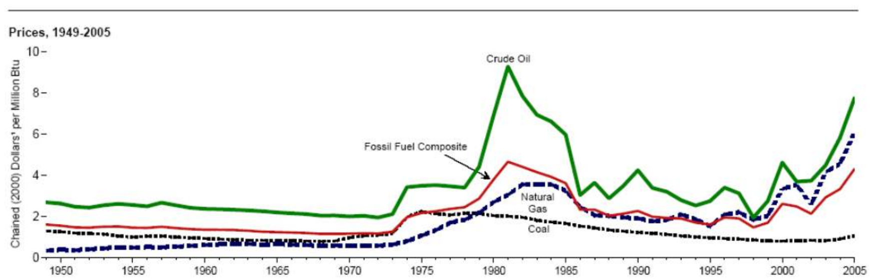

Figure 4.1 Long term fossil fuel cost trends [25]

EIA provides a monthly updated Short Term Energy Outlook, which "industry participants and energy analysts regularly adopt as a 'best estimate' of future energy outcomes"(page 718 in [9]). We use the 2002 data to illustrate the generation of scenarios. Figure 4.2 was released in January 2002 with estimated NG prices for the whole year [22]. Figure 4.3, released in January 2003, shows the actual 2002 NG prices [24]. The rectangle superimposed on the plots shows the range of forecast prices during 2002. Note that the actual price shown in the second graph is not contained in the rectangle, which indicates substantial inaccuracy in the price forecast. So even though the outlook from EIA is a widely used source based on which utilities and others conduct resource planning and modeling studies, there still exists much inaccuracy and uncertainty.

Using EIA forecasts, uncertain NG cost is modeled as a discrete random variable. There are 3 possible values for each period and 11 periods with uncertain prices in the monthly model,

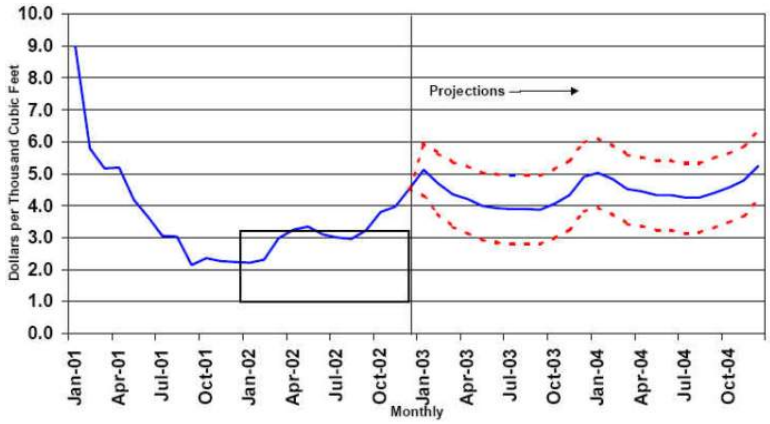

Figure 4.2 EIA short term natural gas price outlooks, Jan. 2002 [22]

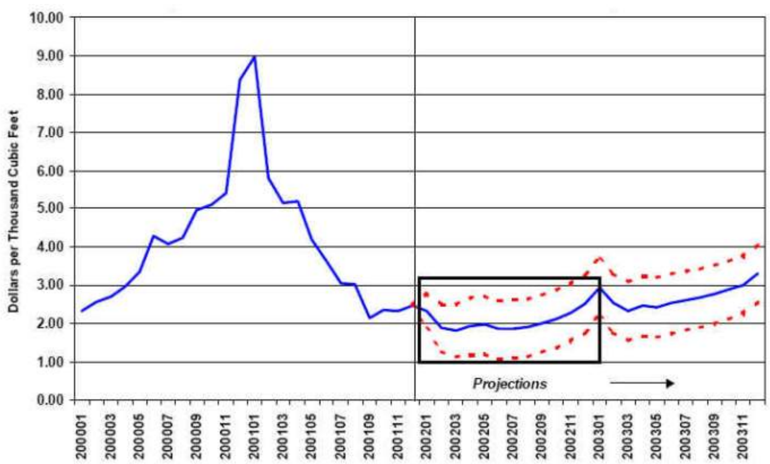

Figure 4.3 EIA short term natural gas price outlooks, Jan. 2003 [24]

assuming the price of January is known. The mean corresponds to the solid lines in figures 4.2 and 4.3. The low value is the lower confidence limit shown in the figure and high value is the upper confidence limit. Both extreme values have the same probability:

$$P\{c_t = LCL_t = \hat{c}_t - CIW_t\} = P\{c_t = UCL_t = \hat{c}_t + CIW_t\} = p_t, P\{c_t = \hat{c}_t\} = 1 - 2p_t.$$

The variance of the random variable $\text{Var}(c_t) = 2p_t(CIW_t)$ depends on p_t and the width of the confidence interval. It is reasonable to set a larger value for p_t for more remote periods because we are more uncertain about the forecast. Case 1 is the base case we will investigate in the next section. Confidence intervals have the constant width shown in figures 4.2 and 4.3.

In order to study the effect of increasing uncertainty, cases 2, 3 and 4 are created by doubling either p_t or CIW_t or both. They will be compared to Case 1 in the next section. The variance of the cost distribution in Case 4 is 8 times that in Case 1.

- Case 1: $p_t = p_{t_0}$, $CIW_t = CIW$ from EIA
- Case 2: $p_t = 2p_{t_0}$, $CIW_t = CIW$ from EIA
- Case 3: $p_t = p_{t_0}$, $CIW_t = 2(CIW$ from EIA)
- Case 4: $p_t = 2p_{t_0}$, $CIW_t = 2(CIW$ from EIA)

Note that, whereas EIA predicts a single national average NG price, we use regional prices in the model. In the original deterministic model, regional prices for each month are generated by multiplying the national price by regional factors derived from the annual data [56]. We assume that future regional prices will have the same relationships to the national average and generate the regional forecasts by multiplying national price forecast by the same factors. Since NG imports from Canada play a very important role in the U.S. national NG consumption, it is necessary to take those NG prices as uncertain elements, too. To generate the forecast for the price of natural gas imported from Canada, we first found the gap between deterministic NG prices in Canada and in U.S.A and then added the difference to the U.S. national NG price forecast.

4.3 Decomposition of the large-scale problem

According to Section 4.2, our 12-month problem can be solved as a sequence of 11 successively smaller two-stage stochastic problems and 1 deterministic problem, among which the largest problem has 3^{11} scenarios. For 2002 data, formulation (3.1) has 1290 nodes ($m_1 = 157$, $m_2 = 952$) and 3480 arcs ($n_1 = 521$, $n_2 = 2959$). Therefore the largest problem written in (3.2) has $157 + 952 \times 3^{11} = 168,644,101$ constraints and $521 + 2959 \times 3^{11} = 524,178,494$ variables, which cannot be solved on a regular PC due to memory limitation.

Benders decomposition [6] and approaches derived from it are one series of schemes that decompose a large size problem into a master problem and several subproblems. The master problem and the subproblems usually iteratively generate bounds that will eventually converge to the optimal solution to the original problem. Problem (3.2) can be decomposed into one master problem and $|S(\hat{t})|$ subproblems using the L-shaped method by Van Slyke and Wets [67]. Through the technique of decomposition, the multi-million variable/constraint problem was solved within the time scale of several days.

CHAPTER 5 RESULTS

5.1 Stochastic model vs. deterministic model for 2002

The model formulated in Chapter 3 and implemented as described in Chapter 4 is solved by three different approaches which lead to three sets of solutions. The true value (TV) solution is obtained from solving the deterministic problem (3.1) with the actual fuel price. The expected value (EV) solution is also from the deterministic problem (3.1) but solved by replacing the actual price with the mean value of its forecast over the whole year. The recourse problem (RP) solution is obtained by solving the stochastic problem via Benders decomposition with the rolling two-stage procedure. The total costs in the last row are the costs encountered if the decisions are implemented in reality under the actual fuel prices.

We first compare the total flows (Table 5.1) in each solution. In the RP solution that contains uncertainty, coal deliveries decrease and NG deliveries increase; especially, imports from Canada are more than doubled relative to the EV solution. As a result, electricity generated from coal-fired plants is reduced and more electricity is generated from natural gas. The electricity trade among regions in RP is less than 90% that in TV. One explanation for the reduction of trade is that, because decision makers could not know the price of NG would soar, they did not buy as much electricity from the areas with cheaper fuel.

Compared to RP, the EV solution is closer to TV, the optimal solution with perfect information. However, RP is closer to the 2002 actual data than either EV or TV is, as shown in Table 5.1. The comparison indicates that while EV and TV rely more on coal, RP has a similar tendency as actually occurred to use more natural gas. In the stochastic case, more natural gas is imported from Canada, which is also closer to reality. We conjecture that this interesting result comes from the greater realism of the stochastic model: we modeled some

Table 5.1 *TV*, *EV*, *RP* solutions compared to 2002 actual data

Results	Actual	*TV*	*EV*	*RP*	(*RP*-*TV*)/*TV*
Coal deliveries (m* tons)	976	1,053	1,049	937	-11.07%
Canada Natural gas deliveries (m Mcf)	886	98	207	617	528.46%
Domestic Natural gas deliveries (m Mcf)	4,785	3,732	3,701	5,337	43.01%
Total Natural gas deliveries (m Mcf)	5,398	3,830	3,908	5,954	55.47%
Electricity generation from coal (m GWh)	1,933	2,117	2,105	1,877	-11.32%
Electricity generation from NG (m GWh)	691	415	426	653	57.50%
Net trade (m GWh)	N/A	381	369	324	-15.12%
Total costs (m $)	N/A	36,668	37,419	42,317	15.41%

* m = million

of the uncertain factors that people making decisions faced in reality. The results indicate that the stochastic model can be utilized as a tool to investigate and predict how the whole system would react under real world uncertainties.

Besides total flows, it is also beneficial to look at regional flows. Figure 5.1 shows that *TV*, *EV* and *RP* make different decisions on how much to buy at each natural gas supply area. The randomization of natural gas cost not only changes the total flows but also has a significant impact on the amount of natural gas purchased from each supply area.

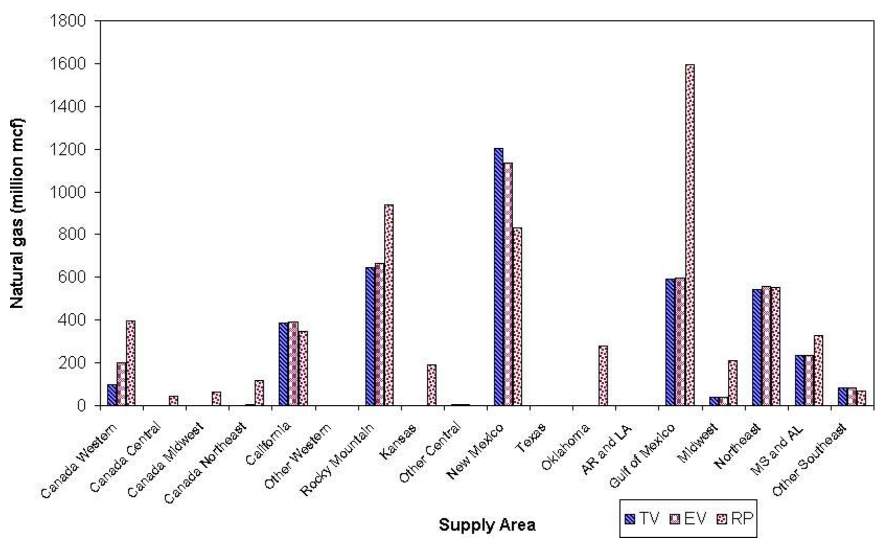

Figure 5.1 Natural gas flows from supply areas, 2002

Murphy and Sen [46] showed that the optimal solution to a stochastic linear program has at least as many nonzero values for first-stage variables as does the optimal solution to the deterministic problem solved for any of its scenarios. Thus, the stochastic solution tends to be more diversified than the deterministic solution. In our problem, TV has 2687 basic variables with nonzero values, EV has 2713 and RP has 2749. Our results agree with the Murphy and Sen's conclusion in two ways: (1) RP is less concentrated on coal; (2) natural gas is supplied from sixteen regions in RP, which is four more than those in TV, as shown in Figure 5.1.

Natural gas storage levels in TV, EV and RP are compared in Figure 5.2 with the solid line showing the forecasted price trend. When uncertainty is introduced, the system stores more natural gas. Moreover, the storage level in RP is more consistent with the price outlook than those in TV and EV. Figure 5.3 shows the net trade amount at each electricity demand center. At most locations, exports or imports decline because of future price uncertainty, which corresponds to the decrease of total power trade in the total flows comparison (Table 5.1).

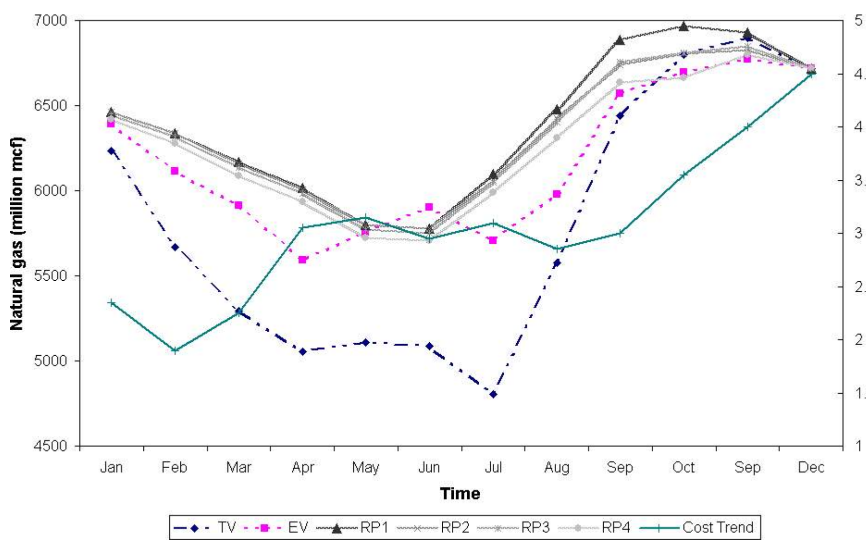

Figure 5.2 Natural gas storage levels, 2002

5.2 Stability of the model

In a stochastic model, scenarios represent users' subjective views on how the real situation is best represented by the data. A stochastic program that provides very different first stage decisions with respect to changes in the underlying probability measure is inconsistent and

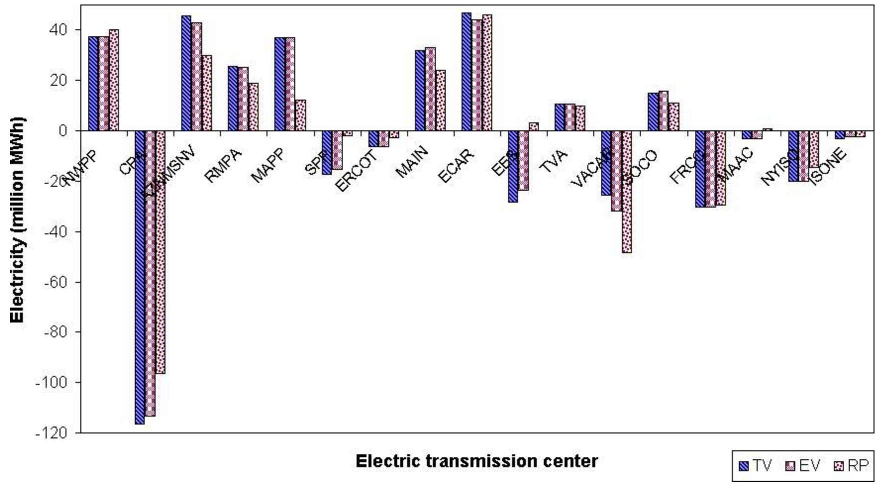

Figure 5.3 Electricity exports at demand centers, 2002

unstable [58]. The stability of the model in this paper is tested by studying the impact of the degree of uncertainty on the RP solution. We increased the variance of the random variables by changing confidence interval widths and associated probabilities. Case 1 is the benchmark case used in previous analysis. Cases 2, 3 and 4 are as described in Section 4.2. $RP1$ to $RP4$ correspond to the solutions resulting from Cases 1 through 4, respectively.

The solutions of $RP1$, $RP2$, $RP3$, and $RP4$ in Table 5.2 are quite similar to each other in terms of coal delivery, natural gas delivery and the total expected cost. Figure 5.2 shows that they all have higher storage levels than TV, which is apparently an outcome of the need to hedge against future uncertainty by storing more fuel. While converting from the forecast confidence intervals to the discrete distributions was basically guesswork, the similarity among recourse solutions in the four cases indicates that the stochastic solution is stable and not sensitive to the values of p_t and CIW_t that specify the discrete distributions used to generate scenarios. This is consistent with the findings in the small example of Section 3.3.

5.3 Load decomposition

In light of the results from the recourse model as well as the comparison with actual data, it appears that the deterministic network flow model underestimates the usage of natural gas in favor of coal. The TV solution used 29% less natural gas to generate electricity than was

Table 5.2 2002 TV and RP solutions: Case 1 – Case 4

Results	TV	$RP1$	$RP2$	$RP3$	$RP4$
Coal deliveries (m tons)	1,053	937	931	934	931
Canada Natural gas deliveries (m Mcf)	98	617	609	613	709
Domestic Natural gas deliveries (m Mcf)	3,732	5,337	5,427	5,390	5,827
Total Natural gas deliveries (m Mcf)	3,830	5,954	5,789	5,857	6,536
Electricity generation from coal (m GWh)	2,117	1,877	1,868	1,872	1,822
Electricity generation from NG (m GWh)	415	653	662	658	708
Net trade (m GWh)	381	324	324	324	315
Total costs (m $)	36,668	42,317	42,565	42,463	44,582

actually used in 2002. Another possible explanation for the model's underemphasis on natural gas is its aggregation of electricity demand over long time periods. Most of the generating units employed to satisfy peak demand are gas-fired, but the aggregated model might not capture the need for using them because it ignores the daily/hourly variation in load. To test whether some of the difference in NG consumption levels between the deterministic model and the actual data was caused by load aggregation, the electricity load was disaggregated with respect to time in the TV model according to a load duration curve (LDC).

The LDC arranges the demand data in decreasing order of magnitude, rather than chronologically. As most of the publicly available information of load consists of hourly data, aggregating the similar hours in LDC would be an appropriate way to account for the demand variability for mid- or long-term planning problems. Because the hourly load data were only available for regions New York (NY-ISO) and New England (ISO-NE), we decomposed the load of every region according to the pattern of NY-ISO, where the demand of electric power is always intensive and the peak hours are especially critical. In the deterministic model, the 744 hours in each month are sorted in decreasing order and then clustered into ten levels with equal time interval. Figure 5.4 illustrates the procedure for NY-ISO in July 2002. The corresponding generation capacity for each level of load is one tenth of the total regional capacity in one period.

Load decomposition raised the output of natural gas-fired power plants in 14 out of the 17 regions. And the total natural gas consumption increased to 4,228 million Mcf, which is 17%

less than the actual consumption (5,398 million Mcf) in 2002. However, we chose to retain aggregated load because (1) hourly load data are not available for most of the regions except NY and NE and the total increase of NG usage due to load decomposition is approximated by employing the LDC of NY for all the regions, which tends to overstate the impact of peak hours by neglecting regional variations in the demand pattern; and (2) with the estimated increment of NG usage, there is still a 17% gap between the TV solution based on load decomposition and the actual usage, while the discrepancy is smaller for the RP solution without load decomposition, which is only 7% lower than the actual usage.

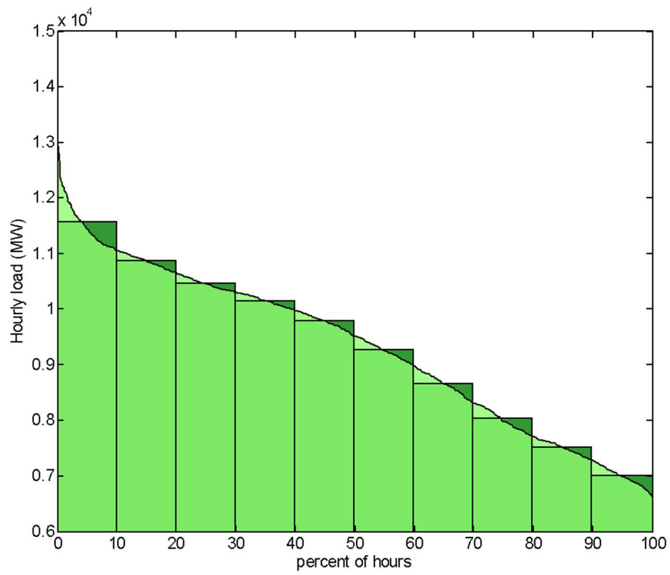

Figure 5.4 Load decomposition using 10 levels

5.4 Results for 2006 data

To test the conclusions drawn from 2002 results, we constructed a 2006 data set, following the same procedure as described in Chapter 4. The most significant difference between the two years' data is that EIA issued outlooks of NG prices that were higher than the actual values in 2006 (figures 5.5 [26] and 5.6 [27]), while for 2002 the forecasts were consistently lower than true prices (figures 4.2 and 4.3). Table 5.3 compares TV, EV and RP for 2006. RP still uses more NG than TV, but they are not as distinct from each other in total consumption of coal and natural gas as in the 2002 results because of the overestimated expense of natural gas. Comparing the 2002 and 2006 results to Case H and Case E, respectively, in section 3.3, reinforces the message that the discrepancy between TV and RP depends on the magnitude and direction of forecast errors.

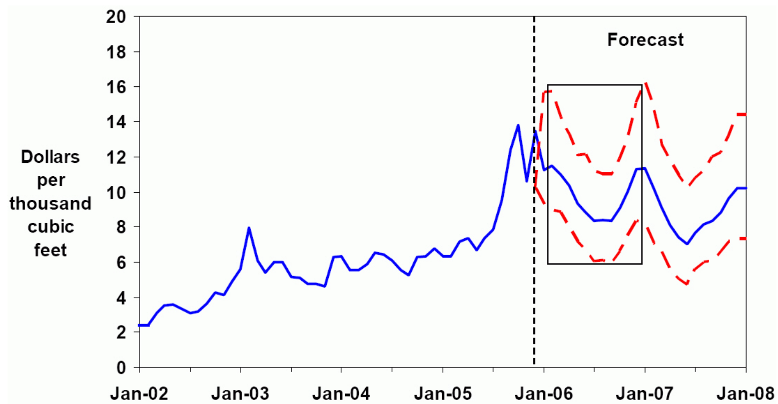

Figure 5.5 EIA short term natural gas price outlooks, Jan. 2006 [26]

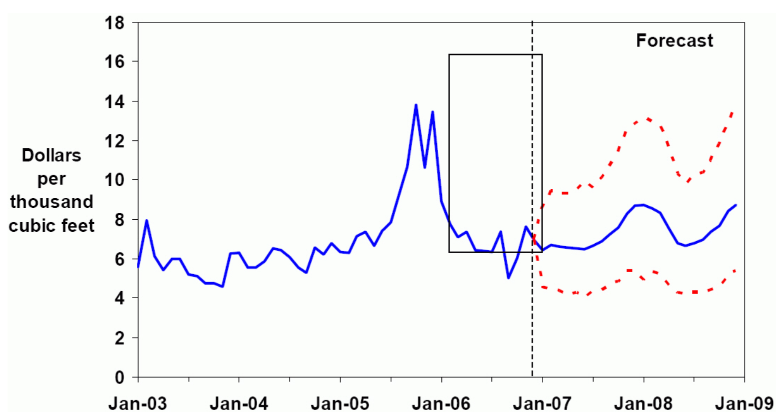

Figure 5.6 EIA short term natural gas price outlooks, Jan. 2007 [27]

Although RP does not differ from TV much in terms of total fuel consumption, introduction of uncertainty leads to differences between the solutions at a more detailed level. The import of NG from Canada is 8.5% lower in the RP solution and it is more similar to 2006 actual data than EV is. Moreover, RP encourages trading electricity because it anticipates a rising trend of NG price. Under this circumstance, importing power is preferred over self-generation.

Table 5.3 TV, EV, RP solutions compared to 2006 actual data

Results	Actual	TV	EV	RP
Coal deliveries (m tons)	1,027	1,082	1,082	1,082
Canada Natural gas deliveries (m Mcf)	933	1,362	1,440	1,251
Domestic Natural gas deliveries (m Mcf)	5,289	3,087	3,020	3,208
Total Natural gas deliveries (m Mcf)	6,222	4,449	4,459	4,458
Electricity generation from coal (m GWh)	1,990	2,141	2,141	2,140
Electricity generation from NG (m GWh)	813	471	471	471
Net trade (m GWh)	N/A	230	236	238
Total costs (m $)	N/A	61,526	63,141	63,081

Figure 5.7 shows that trading activities are increased in 8 out of the 15 electricity demand centers. With respect to diversification, RP has 2428 nonzero basic variables while TV has only 2418. And Figure 5.8 shows that RP purchases NG from two more regions than either TV or EV. RP stores more natural gas than TV and EV as a result of uncertainty (Figure 5.9), and storage is consistent with the cost trend. Finally, a similar set of 4 cases for the distributions is also studied for 2006. Results in Table 5.4 support the conclusion from the 2002 case that the RP solution is stable with respect to the specification of the discrete distributions of random variables. The overestimated NG costs reduce the tendency of more NG consumption in 2006; however, the results remain consistent with those for 2002 in diversification of fuel supply and the impacts of uncertain NG costs on electricity trade and fuel storage.

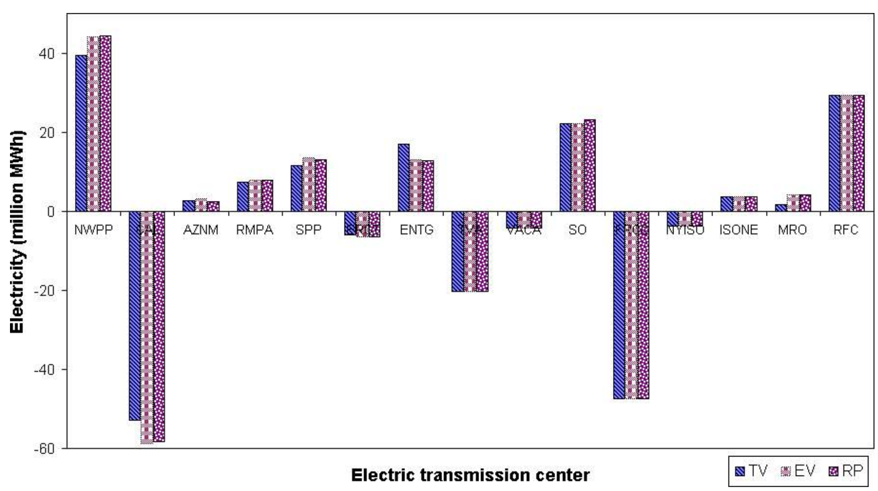

Figure 5.7 Electricity exports at demand centers, 2006

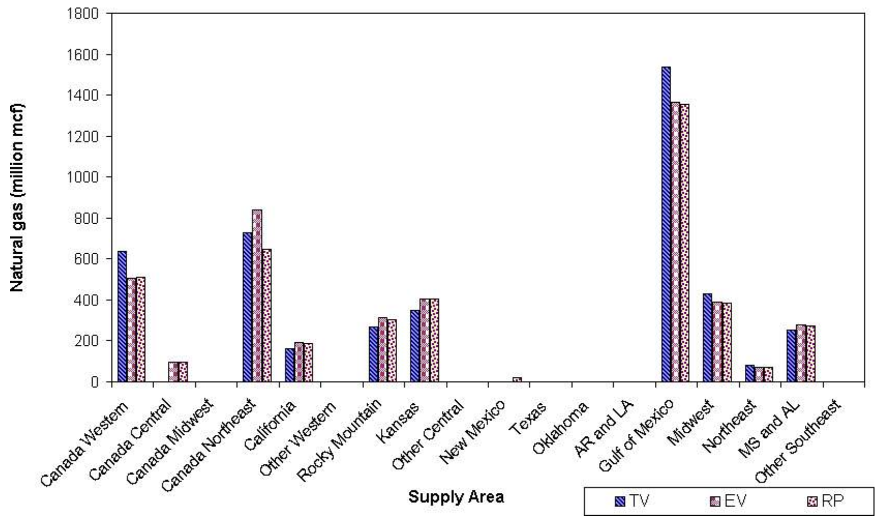

Figure 5.8 Natural gas flows from supply areas, 2006

Table 5.4 2006 TV and RP solutions: Case 1 – Case 4

Results	TV	$RP1$	$RP2$	$RP3$	$RP4$
Coal deliveries (m tons)	1,082	1,082	1,082	1,082	1,082
Canada Natural gas deliveries (m Mcf)	1,362	1,251	1,273	1,259	1,244
Domestic Natural gas deliveries (m Mcf)	3,087	3,008	3,176	3,183	3,010
Total Natural gas deliveries (m Mcf)	4,449	4,458	4,450	4,443	4,253
Electricity generation from coal (m GWh)	2,141	2,140	2,141	2,141	2,141
Electricity generation from NG (m GWh)	471	471	471	471	471
Net trade (m GWh)	230	238	237	240	240
Total costs (m $)	61,526	63,081	62,617	62,511	62,659

5.5 Summary

Our results suggest that, because the stochastic model accounts for the underlying uncertain factors that exist when actual fuel procurement and energy generation decisions are made, the generation mix under stochastic costs is more like the actual situation than the deterministic case (where the differences are more pronounced in the 2002 case). Thus, the stochastic network flow model can be adopted to forecast the actual situation that happens in reality. In a more detailed sense, while coal flows are stable with uncertain NG costs, decisions on natural gas flows vary considerably; in particular, imports from Canada are especially sensitive to cost uncertainty. In addition, more natural gas is stored when the cost uncertainty is considered and

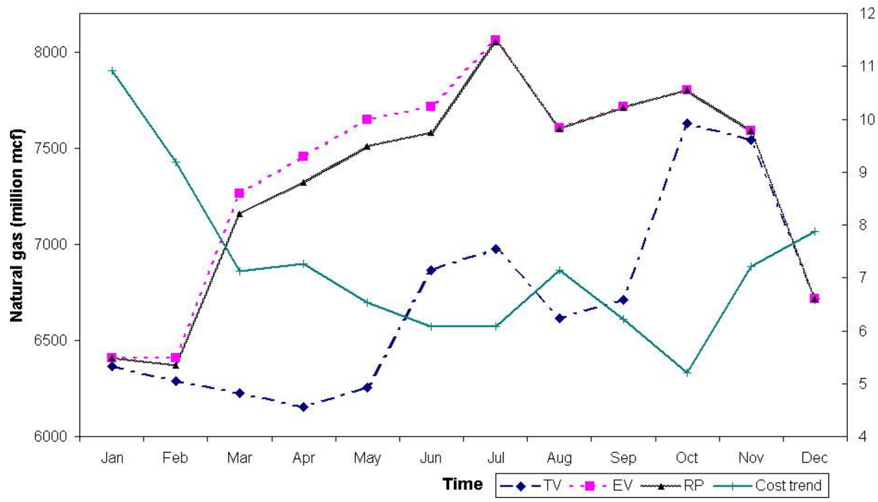

Figure 5.9 Natural gas storage levels, 2006

power trade is highly affected by the outlooks of fuel prices. Finally, the stochastic solutions are consistent under different discrete distributions and thus are robust to errors in the estimated parameters of discrete distributions of the uncertain costs.

CHAPTER 6 METHODS TO ENHANCE COMPUTATIONAL EFFICIENCY

6.1 General scenario reduction

In Section 5.2, we further investigated the effects of uncertainty by varying the distributions of the forecasts. For each distribution, RP must be re-computed. However, it takes several days (5 days for the 2002 problem, 3 days for 2006) to find the exact RP solution via Benders decomposition, which is obviously inconvenient for the further study. Therefore, we tried several methods to reduce the problem size and accelerate the speed of computation. The results presented in this chapter have appeared in our paper [70].

We first employed problem-independent algorithms. Dupačová et al. [21] proposed an approach using a probability metric to measure the distance between the original probability distribution of the scenarios and the probability distribution of the reduced scenarios and furthermore identify a near-optimal subset of scenarios given certain cardinality. Two heuristic algorithms are derived from the extreme examples for $|J| = 1$ and $|J| = |S| - 1$, where J is the set of scenarios to be deleted and $|S|$ is the original set of scenarios. Suppose there are k scenarios we want to delete from the original set. The backward reduction algorithm deletes one scenario every time and redoes it for k iterations, each time in line with the condition $|J| = 1$. Similarly, the forward selection algorithm selects one scenario recursively under the condition $|J| = |s| - 1$, where s is the current set of scenarios. Dupačová et al. [21] proved that the heuristic algorithms provide lower and upper bounds for the minimum distance and hence an approximate solution of the exact scenario reduction problem. One of the most important findings is that backward (forward) outperforms forward (backward) when k is less (greater) than $N - k$. In most of the circumstances, scenario reduction is needed when N is very large

Table 6.1 Monthly model: RP_{sr} compared to RP

	2002			2006		
Results	RP	RP_{sr}	$\frac{RP_{sr}-RP}{RP}$	RP	RP_{sr}	$\frac{RP_{sr}-RP}{RP}$
Coal deliveries (m tons)	937	938	0.09%	1,082	1,082	0.00%
Canada Natural gas deliveries (m Mcf)	617	513	-17.00%	1,251	1,255	0.32%
Domestic Natural gas deliveries (m Mcf)	5,337	5,434	1.83%	3,208	3,202	-0.19%
Total Natural gas deliveries (m Mcf)	5,954	5,947	-0.13%	4,458	4,458	-0.01%
Electricity generation from coal (m GWh)	1,877	1,879	0.07%	2,140	2,141	0.00%
Electricity generation from NG (m GWh)	653	652	-0.20%	471	471	-0.03%
Net trade (m GWh)	324	327	1.06%	238	236	-0.80%
Total costs (m $)	42,317	42,078	-0.57%	63,081	63,120	0.06%

and a small set of scenarios are selected, when forward selection is always chosen over backward deletion. Our study on scenario reduction algorithms therefore focus on the former method.

We applied the forward selection algorithm to the monthly model to reduce the maximum number of scenarios from 3^{11} to 20. When the original set of scenario is as large as 3^{11}, it takes nearly 1 hour to select a scenario and the total computation time is around 20 hours, which would be prolonged if we want to include more than 20 scenarios. Table 6.1 shows that the maximum deviations in the outputs of RP_{sr}, which is the RP solution to the 20-scenario SP problem, are 17% for 2002 and .8% for 2006.

6.2 Temporal aggregation

Besides the general algorithm, we also think of the methods which take advantage of the features of the model. Because the large number of scenarios is a result of multiple stages, temporal aggregation was our first choice. Aggregating the monthly model into quarters not only reduces the size of (1) but also reduces the maximum number of scenarios significantly, from 3^{11} to 3^3. The largest deterministic equivalent (3.2) of the quarterly model has $157 + 296 \times 3^3 = 8149$ constraints and $521 + 807 \times 3^3 = 22310$ variables, which is solved in less than 1 second on a 4GB memory PC.

The three sets of solutions to the quarterly models for 2002 and 2006 can be found in tables 6.2 and 6.3. Compared with the RP solution, the quarterly RP solution displays the

Table 6.2 Quarterly model: RP, EV, RP and 2002 actual data

Results	Actual	RP	EV	RP	$(RP\text{-}RP)/RP$
Coal deliveries (m* tons)	976	1,072	1,071	1,018	-5.04%
Canada Natural gas deliveries (m Mcf)	886	119	119	467	292.44%
Domestic Natural gas deliveries (m Mcf)	4,785	3,719	3,719	4,544	22.18%
Total Natural gas deliveries (m Mcf)	5,671	3,839	3,839	5,011	30.53%
Electricity generation from coal (m GWh)	1,933	2,121	2,121	1,997	-5.85%
Electricity generation from NG (m GWh)	691	410	410	533	30.00%
Net trade (m GWh)	NA	350	346	309	-11.71%
Total costs (m $)	NA	35,694	35,996	38,405	7.60%

Table 6.3 Quarterly model: TV, EV, RP and 2006 actual data

Results	Actual	TV	EV	RP	$(RP\text{-}TV)/TV$
Coal deliveries (m* tons)	933	1,458	1,362	1,267	-13.14%
Canada Natural gas deliveries (m Mcf)	5,289	2,876	2,989	3,088	7.39%
Domestic Natural gas deliveries (m Mcf)	6,222	4,334	4,351	4,355	0.48%
Total Natural gas deliveries (m Mcf)	5,671	3,839	3,839	5,011	30.53%
Electricity generation from coal (m GWh)	1,990	2,155	2,155	2,155	0%
Electricity generation from NG (m GWh)	813	456	456	456	0.04%
Net trade (m GWh)	NA	242	253	251	3.66%
Total costs (m $)	NA	61,927	62,416	62,563	1.03%

same trend as the monthly RP solution does. Although the differences between RP and RP are muted due to aggregation, the RP solution does have lower coal deliveries and higher natural gas purchases than RP. The net trade is lower in RP for 2002 and higher for 2006, which is consistent with our findings in Chapter 5. The results validate that the quarterly model is able to show similar effects of uncertain fuel costs on the optimal energy flows and meanwhile reduces the computational load considerably.

6.3 Structural scenario reduction

One drawback of temporal aggregation is that the decision frequency is forced to decrease from monthly to quarterly, which makes the decisions from the quarterly model less usable to decision makers than those from the monthly model. While retaining the monthly data and structure, we exploit the structural features of the problem and the scenarios. section 3.2

Table 6.4 Monthly model: RP' compared to RP

	2002			2006		
Results	RP	RP'	$\frac{RP'-RP}{RP}$	RP	RP'	$\frac{RP'-RP}{RP}$
Coal deliveries (m tons)	937	937	0%	1,082	1,082	0%
Canada Natural gas deliveries (m Mcf)	617	613	-0.65%	1,251	1,251	0%
Domestic Natural gas deliveries (m Mcf)	5,337	5,341	0.07%	3,208	3,207	-0.03%
Total Natural gas deliveries (m Mcf)	5,954	5,954	0%	4,458	4,458	-0.02%
Electricity generation from coal (m GWh)	1,877	1,878	0.05%	2,140	2,141	0%
Electricity generation from NG (m GWh)	653	653	0%	471	471	-0.02%
Net trade (m GWh)	324	326	0.62%	238	238	-0.09%
Total costs (m $)	42,317	42,273	-0.1%	63,081	63,101	0.03%

explained that only first-stage decisions are kept after each two-stage problem is solved. Thus instead of the general distance measurement in the scenario space, it is more important to evaluate the impact of a scenario on the first-stage decisions. Considering the fuel storage which connects two successive periods, yet without strict theoretic proof, we found that the forecasts of nearer periods have greater impact on the first period decisions. Suggested by this structural feature, we cut the number of scenarios by using only the most likely costs in remote periods, as illustrated in Figure 6.1 with a 6-period problem.

For implementation, we retain the extreme cost values in the first 6 periods so that the maximum number of scenarios is reduced to $3^6 = 729$. The RP solution to (3.2) with reduced scenarios is called RP'. It is compared to RP in Table 6.4. RP' is quite close to RP in both 2002 and 2006 with the largest deviation of .65%. Both results are better than RP_{sr} and it takes only 1 hour to solve for RP'. That is much shorter than the approximately 729 hours that would be used by the general algorithm if 729 scenarios are selected.

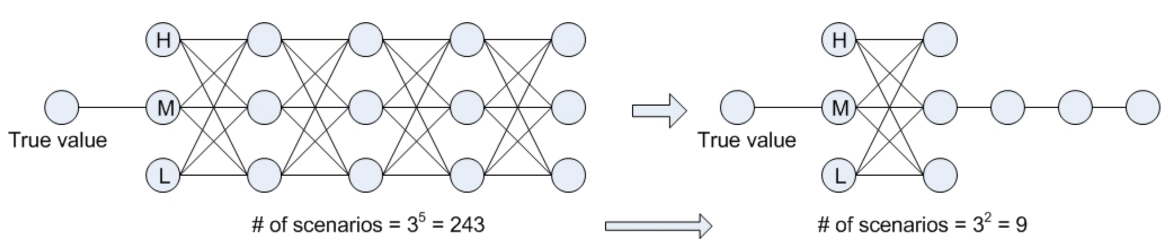

Figure 6.1 Structural scenario reduction: single value for further periods

6.4 Summary

In order to ease the computational load and accelerate calculation speed, three methods are employed and compared. Both temporal aggregation and structural scenario reduction accomplished considerable problem size reduction and maintained accuracy of the conclusions drawn from the RP approximations. Without making use of the model features, the general scenario reduction method costs greater computational efforts. In terms of calculation load, the convergence time for Benders decomposition is problem dependent and highly correlated to the initial point. The heuristic solutions are obtained within relatively stable computation time because the time needed to selecting one scenario is certain once the original set of scenarios are set, while in Benders decomposition, the number of iterations before convergence is uncertain and highly sensitive to the initial point.

Most of the general scenario reduction/sampling methods have their focus on the measurement of the distance between the selected scenario set and the original scenario set. However, in this application, "only the first-stage decisions matter" is a crucial feature that promotes the idea to select a scenario according to how much impact the scenario has on the first stage. The structural scenario reduction method outperformed the general algorithm in both accuracy and efficiency, which inspired our proposed study of algorithms in the aspect of the accuracy of first-stage decision. In the next chapter, we propose an algorithm that is much more efficient in selecting scenarios than FS and a heuristic method that selects a scenario basing on its impact on the first-stage decisions. The new approaches can be utilized in large-scale applications emphasizing the current period decisions under a multiperiod structure.

CHAPTER 7 HEURISTIC ALGORITHMS FOR SCENARIO REDUCTION

7.1 Restrictions of the general scenario reduction method

A general stochastic programming problem can be written as

$$v(P) = \min_{x \in X} E_P f(x, \omega) = \int_{\Omega} f(\omega, x) P(d\omega), \tag{7.1}$$

where ω is the set of random parameter(s) and P is the original probability measure of ω. In parallel to formulation (3.2), $E_P f(x, \omega)$ is the total expected cost $\sum_{s \in S(\hat{t})} \pi_s f_s(z, y^s)$, ω corresponds to the uncertain costs $c_{ij}(t), (i, j; t) \in A_2$, and P is the probability measure of S, the original set of scenarios. The general scenario reduction (GSR) method uses a probability metric $d_{f,\rho}(P, Q)$ to bound the discrepancy between the optimal value $v(P)$ and $v(Q)$, the optimal value under the reduced probability measure Q. The probability metric is then further bounded by the distance between Q and P. There exists a constant $\rho > 0$ such that [21]:

$$|v(P) - v(Q)| \leq d_{f,\rho}(P, Q) \leq g(\rho) \hat{\mu}_c(P, Q). \tag{7.2}$$

Here $d_{f,\rho}(P, Q) := \sup_{x \in X \bigcap \rho B} |\int_{\Omega} f(\omega, x) P(d\omega) - \int_{\Omega} f(\omega, x) Q(d\omega)|$, where ρB is a Borel σ-field with radius of ρ. The probability metric $d_{f,\rho}(P, Q)$ is the maximum distance between the two integrals for all the functions f satisfying that $x \mapsto E_Q f(\omega, x)$ is lower semicontinuous, proper and convex, and $g(\rho)$ is a nondecreasing function. The quantity $\hat{\mu}_c(P, Q)$ is the distance between probability measures P and Q, evaluated by integrating a continuous symmetric function c over (P, Q) in the space of (Ω, Ω). The choice of c depends on the quantitative continuity properties of the integrand $f(x, \omega)$. According to the research of Heitsch et al. [31] on the stability of stochastic programs, $c(\omega_i, \omega_j) = \|\omega_i - \omega_j\|$ is proper for general two-stage problems, and $c(\omega_i, \omega_j) = \|\omega_i - \omega_j\|^2$ can be used for multi-stage problems if the cost and right-hand-side are random and the technology matrices are deterministic. Inequality (7.2) means that for any approximation Q of P, the largest deviation in estimating the optimal solution to a stochastic programming problem is bounded by the product of the defined distance between P and Q and a nondecreasing positive function; consequently, the error for the specified problem v is no larger than that.

Although the value of ρ changes with $f(\omega, x)$ so that inequality (7.2) holds, the optimal solution Q and the distance $\hat{\mu}_c(P, Q)$ is problem independent. Hence, the first potential pitfall of GSR is that deviation of $v(Q)$ from $v(P)$ can be very large for the problems having a big $g(\rho)$. Moreover, in (7.2), the first-stage variables and recourse solutions to the SP problem are considered as a whole while it is unclear how accurate is the estimation of the first-stage decisions, which are the most crucial part in many applications of SP.

There is also an issue with implementing GSR. While (7.2) holds for general SP problems, our primary concern is to reduce the number of scenarios in the case where ω is discretely distributed. According to [21], when the original probability distribution P is discrete and carried by finitely many scenarios $\omega_i \in \Omega$ with weights $p_i > 0, i = 1, ..., N$ and $\sum_{i=1}^{N} p_i = 1$, the optimal reduction concept described in (7.2) suggests to calculate the optimal distance $\hat{\mu}_c(P, Q)$ between P and Q following the formula:

$$\hat{\mu}_c(P, Q) := \min \{D_J = \sum_{j \in J} p_j \min_{i \notin J} c(\omega_i, \omega_j) : J \subset \{s_1, ..., s_N\} = S, |J| = k, N - k = n\}, \quad (7.3)$$

where J is the set of deleted scenarios and $\omega_i \in S \setminus J$ are the retained scenarios. The distance between a deleted scenario ω_j and the retained set equals the minimum distance between ω_j and $\omega_i, i \notin J$, measured with function $c(\omega_i, \omega_j)$. Given the cardinality of J, the goal is to delete the k scenarios whose distances with the retained set sum to the minimum, as described in (7.3). The retained set of scenarios is Q with $\omega_i, i \notin J$ and $q_i = p_i + \sum_{j \in J_i} p_j$ where $J_i := \{j \in J : i = i(j)\}$ and $i(j) \in \arg\min_{i \notin J} c(\omega_i, \omega_j)$ for each $j \in J$. The coefficients q_i obtained in this way are called the optimal weights.

There are $\binom{N}{k}$ ways to choose k scenarios from a set of N. A scenario reduction heuristic is usually needed when N is very large, in which case it is hopeless to solve the special combinatorial optimization problem (7.3) by enumerating the feasible solution set. However, Dupačová et al. [21] developed the two heuristic algorithms of Forward Selection (FS) and Backward Reduction (BR), which are based on the extreme cases $k = N - 1$ and $k = 1$, respectively. FS/BR selects/deletes scenarios recursively and the computation load increases with the number of original scenarios. The most time consuming step is the calculation of $c(\omega_i, \omega_j)$. We present the iterative algorithms in Tables 7.1 and 7.2 and measure the computational efforts in terms of the number of times that $c(\omega_i, \omega_j)$ is calculated. In FS, the first scenario is selected such that the sum of its distances to the unselected scenarios is minimized. Then in each iteration, a new scenario is selected such that by adding this scenario to the selected set, the reduction of the distance between the selected scenarios and the deleted scenarios is maximized. This is done by putting each unselected scenario into Q tentatively, recording the new distance between P and Q and choosing the scenario resulting in the smallest new distance. The algorithm terminates when n scenarios are selected. If t is the time needed to calculate $c(\omega_i, \omega_j)$ for one pair (ω_i, ω_j), when the number of scenarios in N is very large, it takes about $N^2 t$ time to select one scenario. For the problem studied in the previous chapters, N could be as large as $3^{11} = 177,147$, so the computation loads for FS are huge. In BR, the first scenario is deleted such that it has the minimum distance to the remaining scenarios, which means it can be well represented even if deleted. In each iteration, the distribution of the undeleted scenarios is updated by adding the probability of the last deleted scenario to that of its closest undeleted scenario; after the redistribution, another scenario is deleted according to the same criteria by which the first one is deleted. It takes about $0.5N^2 t$ to delete one scenario if N is sufficiently large. Even though FS and BR are able to provide approximations for the intractable problem (7.3), they are not very efficient when N is large, which is usually the case because otherwise reduction of scenarios would not be necessary.

GSR is an optimal scenario reduction method based on stability analysis of the probability metrics, with two heuristics, FS and BR. Theoretically, the reduction of scenarios is problem independent and it could lead to a bad approximation of the exact solution, particularly a

Table 7.1 Forward selection algorithm

Step 1.	Set $i = 1$, $u_i = \arg\min_{u \in S}\{\sum_{j \neq u} p_j c(\omega_j, \omega_u)\}$; $m_j = u_1, \forall j \in S \setminus \{u_1\}$.
Step 2.	$i = i + 1$, $u_i = \arg\min_{u \in S\setminus\{u_1,...,u_{i-1}\}}\{\sum_{j \notin S\setminus\{u_1,...,u_{i-1},u\}} p_j \min_{l \in \{u_1,...,u_{i-1},u\}} c(\omega_j, \omega_l)\}$; if $c(\omega_j, \omega_{u_i}) < d_j$, $m_j = u_i, d_j = c(\omega_j, \omega_{u_i})$, for each $j \in S \setminus \{u_1, ..., u_i\}$.
Step 3.	if $i < n$, go to Step 2; otherwise, calculate the optimal weights $q_{u_i} = \sum_{j:m_j=u_i} p_j, i = 1, 2, ..., n$.
Complexity: $N(N-1) + ... + (N-(n-1)))(N-n) = O(nN^2)$, if $N \gg n$	

Table 7.2 Backward reduction algorithm

Step 1.	Set $j = 1$, $u_j = \arg\min_{u \in S}\{p_u \min_{i \in S\setminus\{u\}} c(\omega_i, \omega_u)\}$; $m_{u_j} = \arg\min_{i \in S\setminus\{u_j\}}\{c(\omega_i, \omega_{u_j})\}$, $p_{m_{u_j}} = p_{m_{u_j}} + p_{u_j}$.
Step 2.	$j = j + 1$, $u_j = \arg\min_{u \in S\setminus\{u_1,...,u_{j-1}\}}\{p_u \min_{i \in S\setminus\{u_1,...,u_{j-1},u\}} c(\omega_i, \omega_u)\}$; $m_{u_j} = \arg\min_{i \in S\setminus\{u_j\}}\{c(\omega_i, \omega_{u_j})\}$, $p_{m_{u_j}} = p_{m_{u_j}} + p_{u_j}$.
Step 3.	if $j < k$, go to Step 2; otherwise, calculate the optimal weights $q_{u_i} = \sum_{j:m_j=u_i} p_j, i = 1, 2, ..., n$.
Complexity: $0.5[N(N-1) + ... + ((N-(N-(k+1)))(N-(N-k))] = O(0.5kN^2)$, if $N \gg k$	

bad first-stage decision. Practically, the heuristic algorithms could be inefficient given a large number of scenarios. In Section 7.2, we develop a heuristic method based on FS in order to lighten the computational burden. In Section 7.3, we address how to better estimate the first-stage solution and propose a new heuristic, which clusters the scenarios according to their individual first-stage solutions and then conducts a selection procedure within each cluster. The proposed algorithms are tested and compared to FS in Section 7.4 on a three-period problem, in Section 7.5 on a regional realization of model (3.2), and in Section 7.6 on the case studies of Chapter 5.

7.2 The accelerated forward selection algorithm

In FS, the scenarios are selected one by one. In each iteration, a scenario is selected and the distance between J and $S \setminus J$ is reduced in two ways. Firstly, the distance between the newly selected scenario and $S \setminus J$ declines to zero from some positive value; secondly, the distances between unselected scenarios and $S \setminus J$ may be decreased because some of them are closer to the newly selected scenario than any of the previously selected scenarios. In FS, the

Table 7.3 Accelerated forward selection algorithm

Step 1.	Set $i = 1$, $u_i = \arg\min_{u \in S}\{\sum_{j \neq u} p_j c(\omega_j, \omega_u)\}$; $m_j = u_1, d_j = c(\omega_j, \omega_{u_1}), j \in S \setminus \{u_1\}$.
Step 2.	$i = i + 1$, $u_i = \arg\max_{u \in S \setminus \{u_1,...,u_{i-1}\}}\{p_u d_u)\}$; if $c(\omega_j, \omega_{u_i}) < d_j$,$m_j = u_i, d_j = c(\omega_j, \omega_{u_i})$, for each $j \in S \setminus \{u_1, ..., u_i\}$.
Step 3.	if $i < \|N - k\|$, go to Step 2; otherwise, calculate the optimal weights q_i.
Complexity:	$N(N-1) + (N-1) + (N-2) + ... + (k+1) = 1.5(N^2 - N) - 0.5k^2$ $= O(N^2)$, if $N \gg (N - k)$

impacts of a newly selected scenario include both ways of distance reduction. When we apply FS in Section 6.1, it is observed that the former impact is dominant. Hence, we propose an accelerated forward selection (AFS) algorithm that takes into account only the first impact as selection criterion. It is presented in Table 7.3. Note that it differs from FS only in step 2. The time needed to calculate $c(\omega_i, \omega_j)$ in AFS is $\frac{1}{n}$ of that in FS. The total computation time saved is much greater than that because HFS bypasses many other operations such as comparisons and summations. Table 7.4 compares HFS to FS by applying it to the 2002 case of (3.2). The scenarios are arranged in certain order and hence identified by the sequence numbers. AFS selected an identical set of scenarios for the largest original set. The maximum discrepancy is two out of twenty scenarios and that happened for the smallest original set.

The reason why the result from AFS is surprisingly consistent with that from FS in this problem is that the scenarios in this problem are close to the average scenario and distant from each other. As illustrated in Figure 7.1, after the center point selected, the triangle point is selected. But including this new point in the selected set will not change the distances from other unselected points to the selected set because the closest point for them is still the center. Accordingly, the first factor of distance reduction is dominant, as observed. Neverthless, the probability distribution of the scenarios can be quite different from this setting. For example, suppose the scenarios are evenly distributed within an interval and the average scenario is in the middle as shown in Figure 7.2, where seven points are sampled from a uniform distribution. The center point is the first selection. Both FS and AFS select the triangle point or the diamond point as the second scenario, although the distance reduction from the second factor

Table 7.4 Comparison of scenarios selected by AFS and FS in 2002 case. Scenarios that differ between the two methods are italicized.

	$\lvert N\rvert = 3^{11}$		$\lvert N\rvert = 3^{9}$		$\lvert N\rvert = 3^{7}$	
Order of selection	FS	AFS	FS	AFS	FS	AFS
1	0	0	0	0	0	0
2	3	6	3	3	3	3
3	6	3	6	6	6	6
4	27	27	2	2	1	1
5	54	54	1	1	2	2
6	9	18	9	9	9	9
7	18	9	18	18	18	18
8	1	2	54	27	27	27
9	2	1	27	54	54	54
10	486	486	243	81	81	81
11	243	243	486	162	162	162
12	4374	2187	162	243	243	243
13	2187	4374	81	486	486	486
14	81	729	729	729	729	729
15	162	1458	1458	1458	1458	1458
16	729	81	13122	6561	4	4
17	1458	162	6561	13122	8	5
18	6561	13122	2187	2187	*12*	*7*
19	13122	6561	4374	4374	*24*	8
20	19683	19683	*5*	*4*	5	*19*

is 2 and not negligible comparing to 3 from the first factor. Even in the situation where reduction caused by the second factor is relatively large, like this case, selections made by AFS are consistent with those by FS.

Here is an informal justification of AFS's consistency with FS. AFS selects the scenario that is furthest from the selected set so that reduction of the first factor is maximized. At the same time, for those scenarios which are also far away from the selected set but near to this scenario, a large distance reduction from the second factor can be achieved, although it is not calculated. AFS does not select a scenario that is near to the selected set because reduction from the first factor is small. Given that the scenario is near to the selected set, the intuition is that its distances to other unselected scenarios are close to those of the selected set. Hence, reduction from the second factor can not be very large. Therefore, AFS's selection criterion is consistent with that of FS when we consider both factors. In sections 7.4 and 7.5, we evaluate AFS algorithm, together with the heuristic algorithm that is to be proposed in section 7.3, under distinct settings of uncertain parameter distributions.

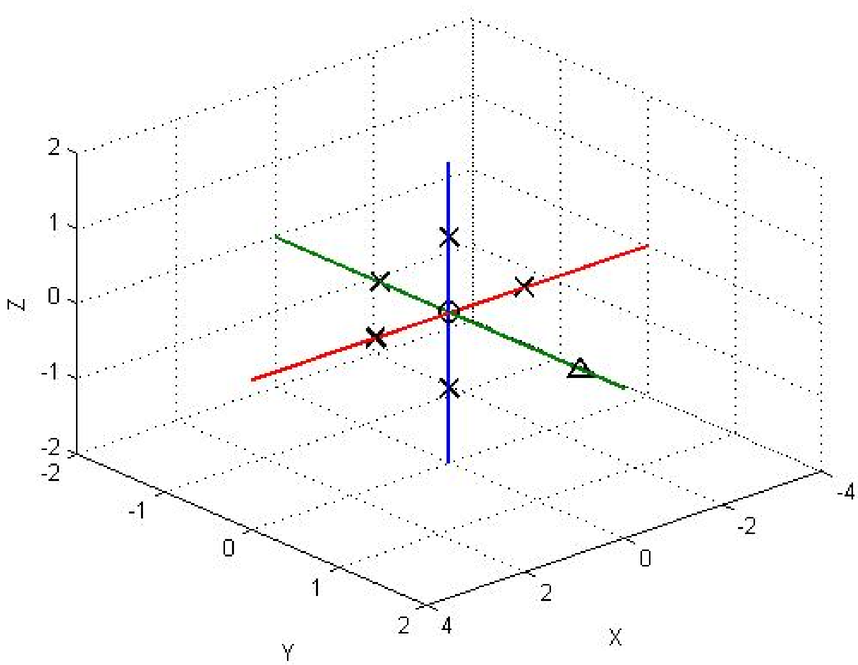

Figure 7.1 A three-dimension illustration of the scenarios in case study 2002

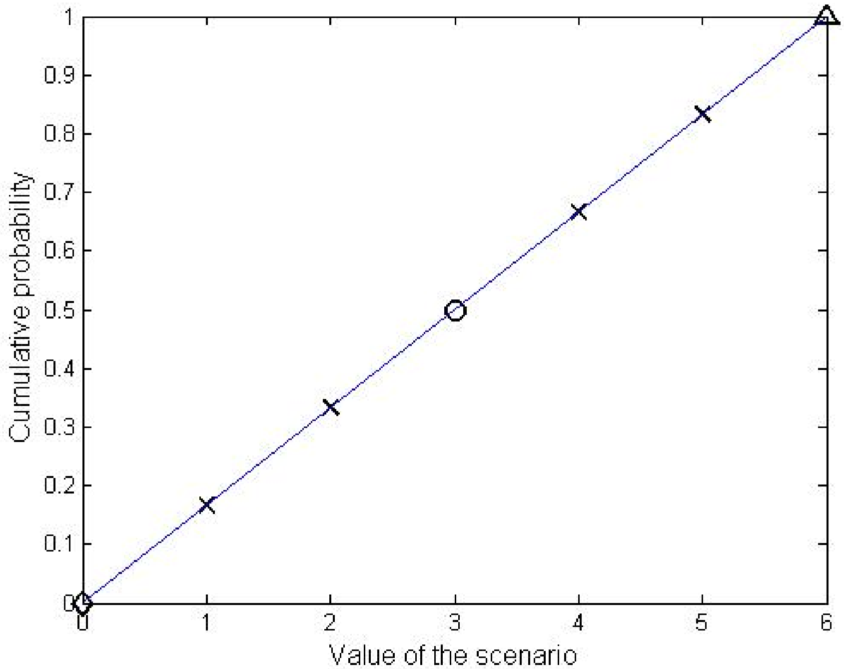

Figure 7.2 A single-dimension illustration of the scenarios sampled from a uniform distribution

7.3 Forward selection within clusters of transferred scenarios

In section 7.2 we devised the accelerated forward selection heuristic that provides a way to improve the efficiency of the scenario reduction method. In this section, we propose another heuristic algorithm where the SP problem is included in the selection procedure. The purpose is to provide tighter bounds on the discrepancy of the first-stage decisions than the general scenario reduction method by making use of the specific program into which the scenarios are incorporated.

Let us go back to look at the combinatorial problem (7.3). Because p_i and ω_i are the only parameters, it accounts for only one aspect of the stochastic program - the set of scenarios, but not the objective function, decision variables, or where the uncertain quantities appear in the formulation. Given a set of scenarios, the selection result is the same no matter what kind of SP is utilizing the scenarios and no matter whether the uncertain parameters are cost

Table 7.5 Forward selection within clusters of transferred scenarios (TCFS)

Step 1.	for each scenario, solve the deterministic problem and keep the key first-stage decisions as the transferred scenario corresponding to the original scenario;
Step 2.	group the original scenarios with the same transferred scenarios; if the number of groups n_g is less than or equal to n, the designated cardinality of the set of selected scenarios, go to step 4;
Step 3.	cluster the n_g transferred scenarios into n clusters using k-means* method and cluster the original scenarios into n groups accordingly;
Step 4.	within each group of the original scenarios, select one scenario using FS algorithm.
	Complexity: $O(N^2)$ assuming each group has N/n scenarios and $N \gg n$

* Implemented with MATLAB built-in function kmeans [MathWorks]

coefficients or right-hand-sides. The measurement of distance could be even problematic when different quantities, such as costs and demands having different numerical magnitudes, are both represented in the scenarios. Therefore, instead of direct selection based on distances $c(\omega_i, \omega_j)$ among scenarios, we propose the algorithm of selection within groups of scenarios, which are clustered according to their impact on the first-stage decisions. A detailed description of the method is in Table 7.5.

For a large-scale SP problem, the number of first-stage variables is usually large, too. It is difficult to measure the impact of a scenario if all these variables are tracked. Thus, in Step 1, we suggest that only key first-stage variables be included in the transferred scenario. Key variables are the variables which determine others. For example, fuel purchase at a power plant decides the amount of electricity generated by the plant, therefore, the former variable, determining the latter one, is a key decison that should be included. Problem 7.4 is a numeric realization of the two-period small example in Section 3.3. Because the first period coal price $c_{11} = 4$ is less than natural gas price $c_{21} = 4.5$, it is straight forward that $x_{11} = 60$ and $x_{21} = 40$. The amount of storage x_{2s} depends on the values of the uncertain future fuel prices and demands. As it is closely related to the scenarios, this variable should be considered as the key first-stage decision to quantify the impact of each scenario.

$$\min_{x} \quad 4x_{11} + 4.5x_{21} + 5x_{2s} + \sum_{s\in S}(c^s_{12}x^s_{12} + c^s_{22}x^s_{22})$$

$$
\begin{aligned}
\text{subject to} \quad & x_{11} \leq u_1 \\
& x_{20} \leq u_2 \\
& x_{12}^s \leq u_1 && \forall s \in S \\
& x_{22}^s \leq u_2 && \forall s \in S \\
& x_{11} + x_{21} = 100 \\
& x_{20} - x_{21} - x_{2s} = 0 \\
& x_{12}^s + x_{2s} + x_{22}^s = d^s && \forall s \in S \\
& x_{11}, x_{12}^s, x_{22}^s, x_{20}, x_{21}, x_{2s} \geq 0, && \forall s \in S
\end{aligned}
\tag{7.4}
$$

Grouping the scenarios before performing forward selection has two advantages. Firstly, selecting one scenario from each of the n (or n_g if $k_g < k$) groups takes much less time than selecting n scenarios from the original set of N scenarios. The value of n is bounded from above by the the constraint $Timeforselection(method, k) + TimeforsolveSP(k) < T_L$, where T_L is the maximum time that is allowed to solving for RP solution. In Chapter 6, the SP problem with several hundred scenarios can be easily solved in less than one hour but $Timeforselection(FS, k) \approx k$ hours limits the number of selected scenarios to 20. The high efficiency of TCFS greatly increases the number of scenarios that can be included in the reduced stochastic programming problem.

In Table 7.6, we use the choosing-two-scenarios-from-four case to illustrate why TCFS is faster than FS. The second period demand is assumed to be certain in this case. In FS, we first calculate the distance between every two scenarios and select scenario 4 which has the smallest sum of distances D_1. In the second iteration, each unselected scenario is put into the selected set tentatively to calculate the updated sum of distances D_2 and scenario 2 is selected. Finally, for the two unselected scenarios 1 and 4, we find out which selected scenario they are closer to, and decide the new probability distribution, shown in column FS. The procedure of TCFS is simpler. According to the transferred scenarios x_{2s}, the scenarios are divided into two groups. Within the first group, scenario 2 is selected because it has the smallest sum of distances among the three scenarios. The new probability of this scenario is the sum of the probabilities of all the scenarios in this group. Since there is only one scenario in the second group, scenario 4 is selected without further calculation. The two methods result in different selections but the same first-stage decision that is consistent with the solution solved with the full set of scenarios. TCFS is more efficient than FS in that it selects scenarios from smaller groups and bypasses the step of calculating new probabilities.

Table 7.6 The two-period small example: TCFS illustration – case 1

		RP	FS			TCFS		
s	P	$(c_{12}^s, c_{22}^s, d^s)$	D_1	D_2	Selection	x_{2s}	D_1	Selection
1	0.25	$(3, 4.8, 100)$	1.11	0.56	N/A	0	0.84	N/A
2	0.25	$(4.2, 4, 100)$	0.91	0.52	0.5	0	0.61	0.75
3	0.25	$(4.9, 4.7, 100)$	0.98	0.52	N/A	0	0.72	N/A
4	0.25	$(4, 5.2, 100)$	0.83	N/A	0.5	20	0	0.25
	x_{2s}	0			0			0
	Cost	846.5			846.5			846.5

The second advantage of TCFS is that the scenarios with similar impact on the first-stage decisions are put in the same cluster for selection, which avoids the situation that a scenario is represented by a scenario close in value but distinct in impact on the first-stage decisions. Case 2 in Table 7.7 is a good illustration. The first three scenarios all represent the trend that in the second period, both coal and natural gas will be cheaper than using storage, while in the last scenario, natural gas will be more expensive than storage. Similar to case 1, TCFS selects scenarios 2 and 4 to represent the two trends and their probabilities. FS, however, relies only on the distance of the values. It selects scenarios 1 and 4 and considers both unselected scenarios 2 and 3 to be closer to scenario 4. The new probability of scenario 4 is so large that the trend represented by this scenario is magnified, which results in a bad approximation of the first-stage decision. The selected set of scenarios by TCFS is problem dependent and produces more accurate first-stage decisions than FS.

TCFS's advantage of making use of problem information can be even more evident when the uncertain parameters are distinct quantities. In Table 7.8, we add the value of uncertain demand into each scenario. Because all of the possible demands are no greater than the total

Table 7.7 The two-period small example: TCFS illustration – case 2

		RP	FS			TCFS		
s	P	$(c_{12}^s, c_{22}^s, d^s)$	D_1	D_2	Selection	x_{2s}	D_1	Selection
1	0.25	$(2, 4.8, 100)$	1.82	0.56	0.25	0	1.31	N/A
2	0.25	$(4.2, 4, 100)$	1.14	0.76	N/A	0	0.83	0.75
3	0.25	$(4.9, 4.7, 100)$	1.23	0.76	N/A	0	0.97	N/A
4	0.25	$(4, 5.2, 100)$	1.07	N/A	0.75	20	0	0.25
	x_{2s}	0			20			0
	Cost	831.5			836			831.5

capacity in the second period, they have no real impact on the storage decision. TCFS's selection remains the same as that in previous cases. Nevertheless, the values of demand dominate those of prices in calculation of distances. Hence, FS's selection again departs from the trends of the original scenarios. Case 4 in Table 7.9 is an instance where the value of demand d^s has influence on the amount of storage. For the sake of feasibility, x_{2s} should be 5 as long as scenario 4 is included. Here, FS does not select this special scenario and the possibility of high demand in the future is not included in the selected scenarios. TCFS, however, is able to recognize scenario 4 because of its divergent impact on storage and therefore produces a more accurate first-stage decision than FS.

Table 7.8 The two-period small example: TCFS illustration – case 3

		RP	FS			TCFS		
s	P	(c_{12}^s, c_{22}, d^s)	D_1	D_2	Selection	x_{2s}	D_1	Selection
1	0.25	$(3, 4.8, 105)$	89.2	13.1	0.25	0	63.9	N/A
2	0.25	$(4.2, 4, 110)$	38.6	13.3	N/A	0	32.0	0.75
3	0.25	$(4.9, 4.7, 120)$	88.9	31.9	N/A	0	82.4	N/A
4	0.25	$(4, 5.2, 115)$	38.4	N/A	0.75	20	0	0.25
	x_{2s}	0			20			0
	Cost	907			911.5			907

7.4 A three-period problem with uncertain costs and demands

The flower-girl problem [14], an extension of the famous newsboy problem [59], is a multi-period problem where the demands are random in each period and the unsold flowers can be

Table 7.9 The two-period small example: TCFS illustration – case 4

		RP	FS			TCFS		
s	P	(c^s_{12}, c_{22}, d^s)	D_1	D_2	Selection	x_{2s}	D_1	Selection
1	0.25	$(3, 4.8, 105)$	133.0	31.8	0.25	0	32.7	N/A
2	0.25	$(4.2, 4, 110)$	70.0	32.0	N/A	0	13.3	0.75
3	0.25	$(4.9, 4.7, 115)$	57.7	N/A	0.75	0	32.4	N/A
4	0.25	$(4, 5.2, 125)$	182.2	32.4	N/A	20	0	0.25
	x_{2s}	5			0			5
	Cost	915			Infeasible			915

carried over to the next period to be sold at a lower price. Here we compose a three-period problem similar to the flower-girl problem except that both the costs and demands for future periods are random. Problem (7.5) is the formulation. The scenario reduction algorithms, FS, AFS and TCFS, are compared on a numeric stochastic version of the problem, which is called the three-period uncertain costs and demands (TUCD) problem in the following text.

$$\begin{aligned} \min_{x_0} \quad & TC = c_0 x_0 + \tilde{c_1} x_1 + \tilde{c_2} x_2 \\ \text{s.t.} \quad & x_0 \geq d_0 \\ & x_0 + x_1 \geq d_0 + \tilde{d_1} \\ & x_0 + x_1 + x_2 \geq d_0 + \tilde{d_1} + \tilde{d_2} \\ & x_0, x_1, x_2 \geq 0 \end{aligned} \tag{7.5}$$

The parameters c_0 and d_0 for the current period are deterministic. The future prices $\tilde{c_1}$ and $\tilde{c_2}$ and future demands $\tilde{d_1}$ and $\tilde{d_2}$ are random variables following distributions described in Table 7.10. Price $\tilde{c_1}$ is uniformly distributed between 4 and 5. Price $\tilde{c_2}$ is the sum of a uniformly distributed random number and an autoregressive term, involving $\tilde{c_1}$. The demand in each period is impacted by the price of that period. While it is impossible to solve for the analytical optimal solution x_0, we can sample a large number of scenarios from the given distributions and then solve problem (7.6) as a good approximation. Each scenario $s \in S$ is composed of values $(c_1^s, d_1^s, c_2^s, d_2^s)$ that are sequentially sampled from the assumed distributions. As the total number of scenarios $|S|$ approaches infinity, the sampled scenarios approach the exact distributions. And it is expected that the first-stage solution to (7.6) approaches to the exact x_0.

Table 7.10 Numeric assuptions of the TUCD problem

c_0	4.5
d_0	10
$\tilde{c_1}$	$\tilde{c_1} = C_1$
$\tilde{c_1}$	$\tilde{c_1} = C_2 + 0.2(\tilde{c_1} - 4.5)$
$\tilde{d_1}$	$\tilde{d_1} = D_1 - 2(\tilde{c_1} - 4.5)$
$\tilde{d_2}$	$\tilde{d_2} = D_2 - 2(\tilde{c_1} - 5)$
C_1	$C_1 \sim uniform(4, 5)$
C_2	$C_2 \sim uniform(4.5, 5.5)$
D_1	$D_1 \sim uniform(15, 25)$
D_1	$D_2 \sim uniform(18, 38)$

$$
\begin{aligned}
\min_{x_0, x_1^s, x_2^s} \quad & c_0 x_0 + \tfrac{1}{|S|} \textstyle\sum_{s \in S} \left(c_1^s x_1^s + c_2^s x_2^s\right) & \\
\text{s.t.} \quad & x_0 \geq d_0 & \\
& x_0 + x_1^s \geq d_0 + d_1^s & \forall s \in S \\
& x_0 + x_1^s + x_2^s \geq d_0 + d_1^s + d_2^s & \forall s \in S \\
\text{Non-anticipativity constraint:} \quad & x_1^{s_1} = x_1^{s_2} & \forall s_1, s_2 \in S, d_1^{s_1} = d_1^{s_2} \\
& x_0, x_1^s, x_2^s \geq 0 & \forall s \in S
\end{aligned}
\tag{7.6}
$$

In our implementation, 2500 scenarios are sampled to formulate (7.6) and solve for the RP solution. Then we use FS, AFS and TCFS to select 50 scenarios from the original set of 2500 scenarios. The discrete SP (7.6) is composed for each of the reduced set of scenarios and an approximated first-stage solution is obtained. The results from the three algorithms are compared to RP in Table 7.11. TCFS achieved the best accuracy in approximating the first-stage decision, while the estimates from FS and AFS are both reasonably good. The low efficiency is a obvious drawback of FS algorithm. It takes a much longer time to select the scenarios than to solve the original formula with 2500 scenarios. Both of the proposed heuristic algorithms outperform FS in computation time. Accelerated forward selection uses the least total time, while TCFS is the most efficient in selection.

In section 7.2, we learned that the distribution of the scenarios affects performance of a scenario reduction algorithm. Another set of 2500 scenarios are sampled under the assumption that C_1, C_2, D_1 and D_2 follow normal distributions with the same mean value and variance. The results based on the new scenarios are presented in Table 7.12, confirming that compared to FS, TCFS achieves the best tradeoff between speed and accuracy.

Table 7.11 The numerical example 1: RP vs. FS vs. AFS vs. TCFS

	RP	FS	AFS	TCFS
x_0	54.25	52.76	51.61	53.59
Accuracy of x_0	100%	97.3%	95.2%	98.8%
Cost	193.6	193.8	194.18	193.65
Error of cost	0	.2	.58	.05
Transfer Time	NA	NA	NA	23.93s
Selection Time	NA	1211.20s	0.94s	0.64s
Solution Time	183.25s	0.04s	0.04s	0.04s
Total Time	183.25s	1211.24s	12.60s	24.01s

Table 7.12 The numerical example 2: RP vs. FS vs. AFS vs. TCFS

	RP	FS	AFS	TCFS
x_0	56.15	54.98	54.71	55.80
Accuracy	100%	97.9%	97.4%	99.4%
Cost	182.04	183.45	185.52	182.04
Error of cost	0	1.41	3.48	0
Transfer Time	NA	NA	NA	24.00s
Selection Time	NA	1237.90s	0.73s	0.66s
Solution Time	181.07s	0.04s	0.04s	0.04s
Total Time	181.07s	1237.94s	0.77s	24.70s

7.5 A regional realization of the multi-period energy transportation model

In this section, we create a regional model that has a similar structure to the large model studied in Chapter 3. This model, namely the NYNE model, considers the aggregated power demands that are supplied by gas-fired power plants in New York and New England. As shown in Figure 7.3, GA and GB represent domestic and Canadian gas supplies, respectively. The same type of power plants in one region are aggregated into one node: EB for combined cycle and EG for gas steam. There is also a storage facility where natural gas can be stored for future usage. As a multi-period problem, consecutive periods are connected by the stored fuel carried from one period to the next. The objective is to figure out a best fuel purchase and power generation plan that would minimize expected total cost given discrete distribution (scenarios) of future natural gas prices.

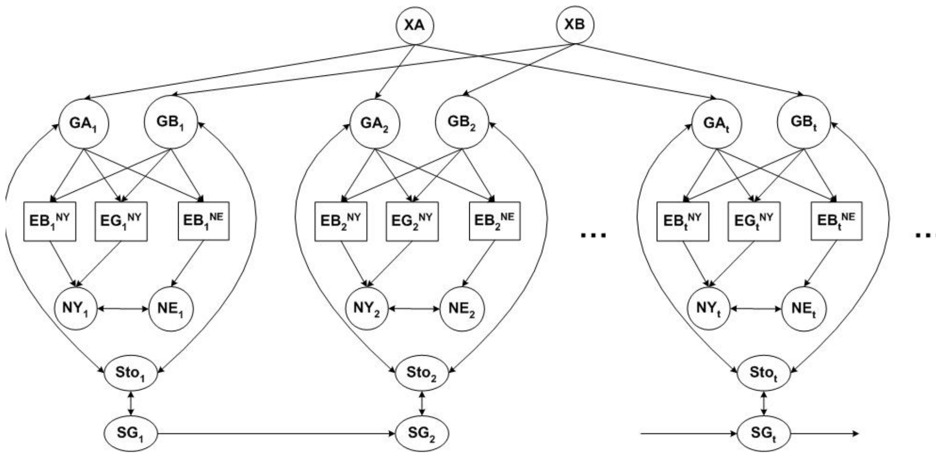

Figure 7.3 A regional realization: the NYNE model

The NYNE model is implemented for year 2008. We collected relevant data from January 2008 to November 2009 so that each two-stage program in the rolling procedure has the same time horizon of 12 months. Natural gas costs and power demands in future periods are independent random parameters, which are assumed to follow normal distributions adopted from the EIA short-term energy outlook. In order to solve for the RP solution, at the beginning of each period, exact NG costs and power demands for the current period are first-stage deterministic parameters and 1000 scenarios are sampled for future uncertainties from the updated forecasts, which are depicted in Figure 7.4. Domestic gas and imported gas have the same mean value for future prices. The variance of Canadian NG prices is twice that for domestic prices. After solving a sequence of 12 two-stage programs, we obtain the RP solution for 2008 case of the NYNE model. The EV solution is also calculated. The two sets of solutions are compared in Figures 7.5 to 7.7. The inclusion of uncertain fuel prices and electric loads has a great impact on the decisions regarding NG imports (Figure 7.6) and storage (Figure 7.7). Consistent with our findings in Chapter 5, more gas is stored in order to hedge the risk associated with future uncertainties. Interestingly, most of the extra storage in the RP solution comes from international purchase that links to higher degree of uncertainty in terms of the variance of future price.

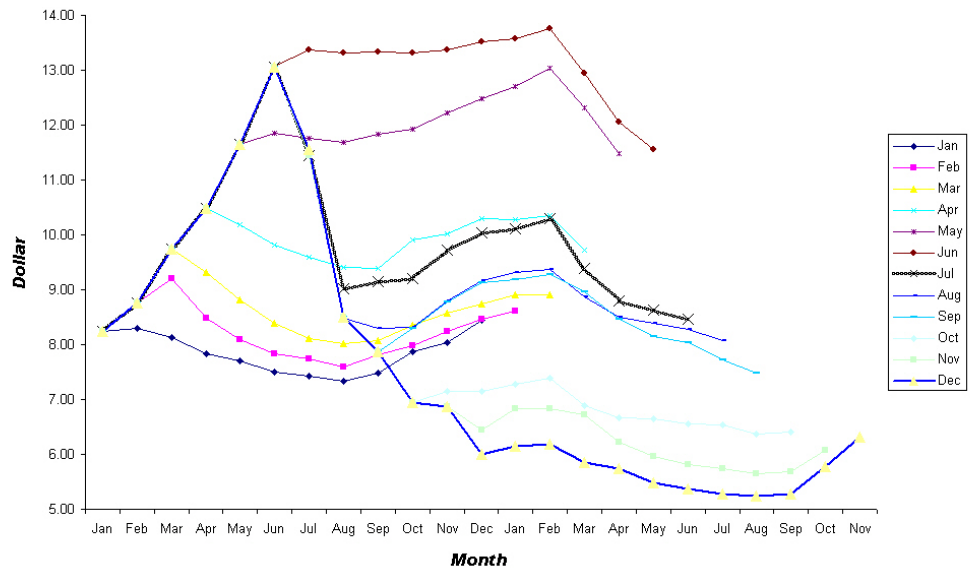

Figure 7.4 Trends of future natural gas prices: 2008-2009

To evaluate the scenario reduction algorithms FS, AFS and TCFS, each is applied to select a subset of scenarios at the beginning of every period and an approximation of the RP solution is obtained by solving the stochastic program composed of the selected scenarios. According to Figures 7.8, 7.9, and 7.10, the domestic purchase of NG does not vary in any of the solutions, while the decisions regarding fuel imports and storage are greatly impacted by selection of scenarios. TCFS produces the solution that is closest to RP. FS and AFS are able to estimate the trend of NG storage. But the two curves are further away from RP than the curve from TCFS.

In Table 7.13, we compare the performances of the three algorithms in detail. Consistent with the graphs, TCFS is the most accurate solution in terms of total absolute deviation from RP. AFS has the highest computational efficiency but its accuracy is not plausible. Neverthless, we can always increase the number of selected scenarios by AFS to reduce the error. FS's low efficiency restricts the size of the selected set and hence the accuracy of the approximation.

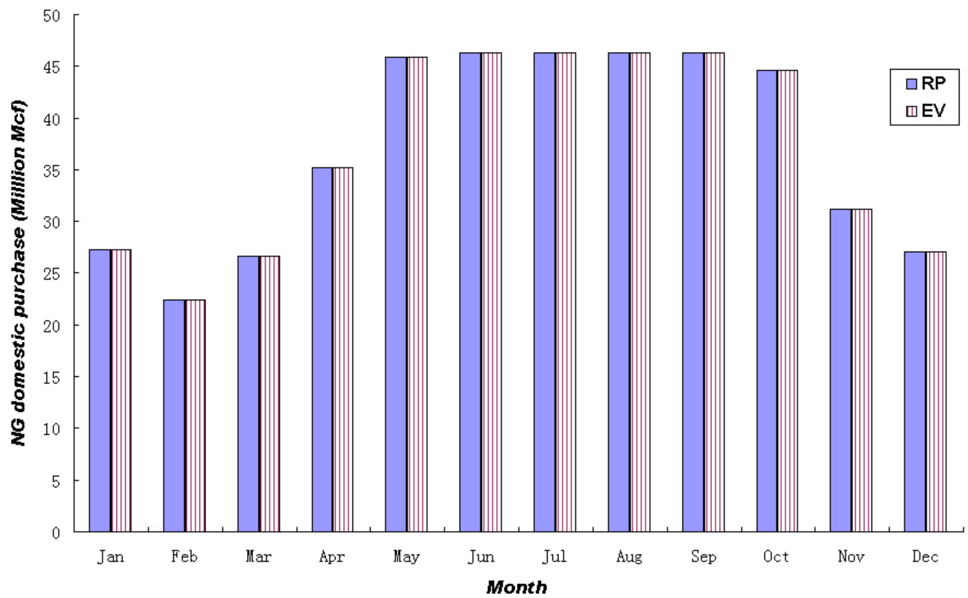

Figure 7.5 The NYNE model 2008 domestic NG purchase: EV vs. RP

TCFS, as we expected, is fast and accurate.

Table 7.13 Comparison of the results from scenario reduction algorithms: the NYNE model

	RP	FS	AFS	TCFS	FS′	TCFS′
# of scenarios	1,000	50	50	25-794*	25-794*	50
Computation time (s)	1,876	3,912	97	502	17,995	577
Cost (million $)	7,527	7,322	7,425	7,433	7,332	7,433
Absolute difference						
Domestic (million Mcf)	0	0	0	0	0	0
Imports (million Mcf)	0	205.11	304.69	141.06	78.31	141.06
Storage (million Mcf)	0	154.55	253.80	123.88	76.08	123.88

*Average = 340; Std = 214.

For both FS and AFS, we have a pre-determined size for the selected set of scenarios. However, when executing TCFS, the value of n is set to be equal to n_g, meaning that we allow a representative scenario for each group that produces a distinct set of key first-stage decisions, which are domestic purchase, import and storage in this problem. The resulting values of n vary from 25 to several hundreds. TCFS is different from AFS and FS in that a larger set of selected scenarios does not mean longer selection time. This is because the more

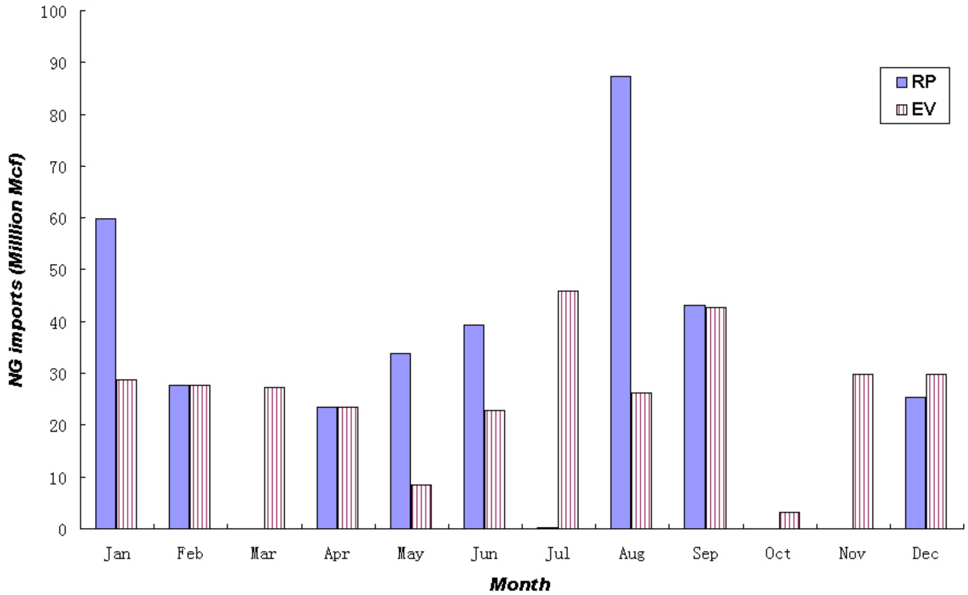

Figure 7.6 The NYNE model 2008 NG imports: EV vs. RP

groups there are, the fewer scenarios in each group and the shorter time is needed to select a scenario from each group. While selection time is not a tight bound in TCFS algorithm, the value of n can still be constrained by the time needed to solve the stochastic program. Plus, it is not clear whether the supreme accuracy of TCFS results from the idea of making use of first-stage decisions or simply from the increased number of scenarios that are selected. In Table 7.13, the column labeled FS′ presents the solution obtained from FS algorithm when n is set to be the same as that in TCFS, while in column TCFS′, n is limited to 50, which is the size in FS and AFS. Figure 7.11 depicts the change of storage trends due to change of the size of selection. Increasing n to a higher level does increase the accuracy of FS algorithm but the huge computational burden greatly weakened the feasibility. With limited n, TCFS yet generates a better solution than FS and AFS, confirming the idea of improving both accuracy and efficiency through utilization of the information provided by the problem itself.

7.6 Comparison of heuristics in case studies of 2002 and 2006

Finally, we applied FS, AFS and TCFS to the complete multi-period energy transportation model implemented in Chapter 4. The results for the 2002 case are presented in Table 7.14. As

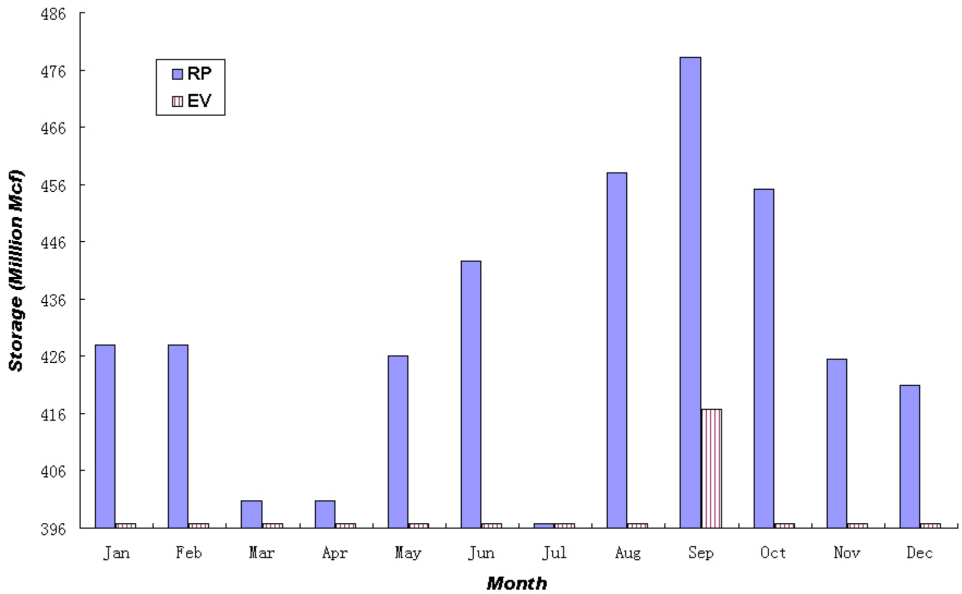

Figure 7.7 The NYNE model 2008 NG storage: EV vs. RP

we discussed in Chapter 6, FS's approximation of total fuel usage is very accurate. However, it does not perform as well in allocating the purchase of NG to imports and domestic supplies. AFS selects very similar subsets of scenarios to FS, as shown in Table 7.4. Hence, the approximation obtained from AFS is very close to that from FS. And AFS achieves the solution in less than one hour, comparing to more than one day used by FS. TCFS, by taking the fuel acquisiton arcs as the key first-stage variables, selects scenarios that have distinct impacts on choices of NG supply locations and thus produces import amount that is much closer to the exact solution RP. The time consumed by TCFS is a little over 3 hours, which is a great improvement in efficiency compared to the time taken by FS.

The difference between FS and RP is reduced in the 2006 case, as shown in Table 7.15. Still, AFS is an efficient way to get a good approximation of the FS results. And TCFS is the best in accuracy.

7.7 Summary

In this chapter, two heuristic scenario reduction algorithms are proposed and demonstrated. With reasoning, we showed that the accelerated forward selection method produces similar

Table 7.14 Comparison of the results from scenario reduction algorithms: 2002 case

Results	RP	FS	AFS	TCFS
# of scenarios selected	3^t	20	20	20
Coal deliveries (m tons)	937	938	938	937
Canada natural gas deliveries (m Mcf)	617	513	512	605
Domestic natural gas deliveries (m Mcf)	5,337	5,434	5,433	5,351
Total natural gas deliveries (m Mcf)	5,954	5,947	5,947	5,956
Total costs (m $)	42,317	42,078	42,078	42,401
Computation time (hour)	436	22.5	0.8	3.6

Table 7.15 Comparison of the results from scenario reduction algorithms: 2006 case

Results	RP	FS	AFS	TCFS
# of scenarios selected	3^t	20	20	20
Coal deliveries (m tons)	1,082	1,082	1,082	1,082
Canada natural gas deliveries (m Mcf)	1,251	1,255	1,255	1,253
Domestic natural gas deliveries (m Mcf)	3,208	3,202	3,202	3,206
Total natural gas deliveries (m Mcf)	4,458	4,458	4,458	4,458
Total costs (m $)	63,081	63,120	63,120	63,101
Computation time (hour)	674	21.7	0.8	3.4

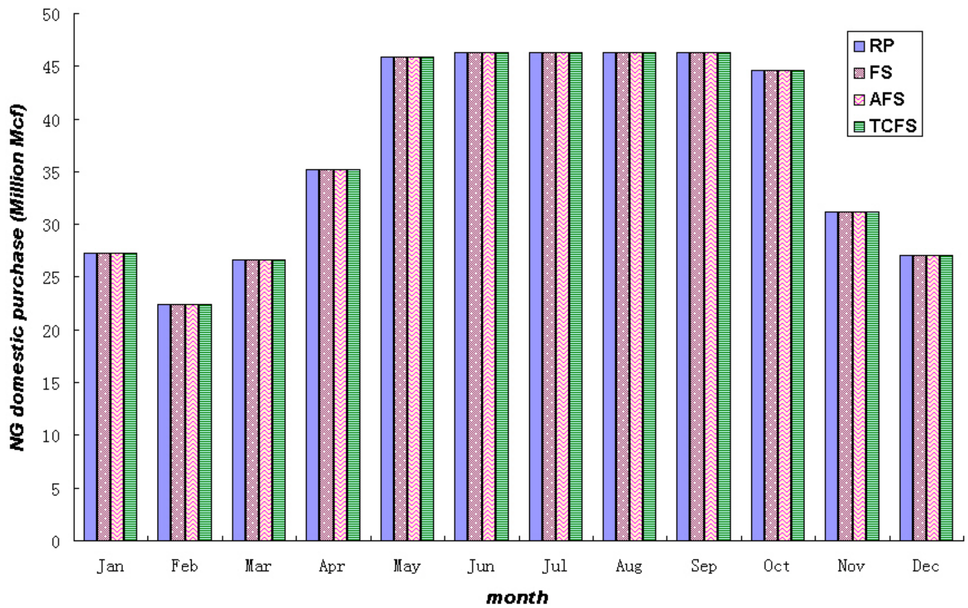

Figure 7.8 The NYNE model 2008 domestic NG purchase: RP vs. FS vs. AFS vs. TCFS

selection to the forward selection method without restiction on scenario distribution and it is much more efficient than FS. If confronted with a large-scale SP problem with a huge number of scenarios, one can apply AFS to select a sub set of scenarios in very short time and get a reasonably good approximation of the exact solution. The second heuristic, TCFS, aims at the two shortcomings of the FS algorithm: it not only reduces the computational burden, but also improves accuracy of the first-stage decisions. By accounting for the impact of each individual scenario on key first-stage decisions, this method divides the scenarios into smaller groups so that the time needed for each selection is greatly reduced. The clustering of scenarios prevents the situation that a scenario is represented by another scenario having distinct impact on the solution. TCFS also avoids the confusing selection results by FS when multiple quantities appear in one scenario and the scale of one quantity dominates in the distance calculation.

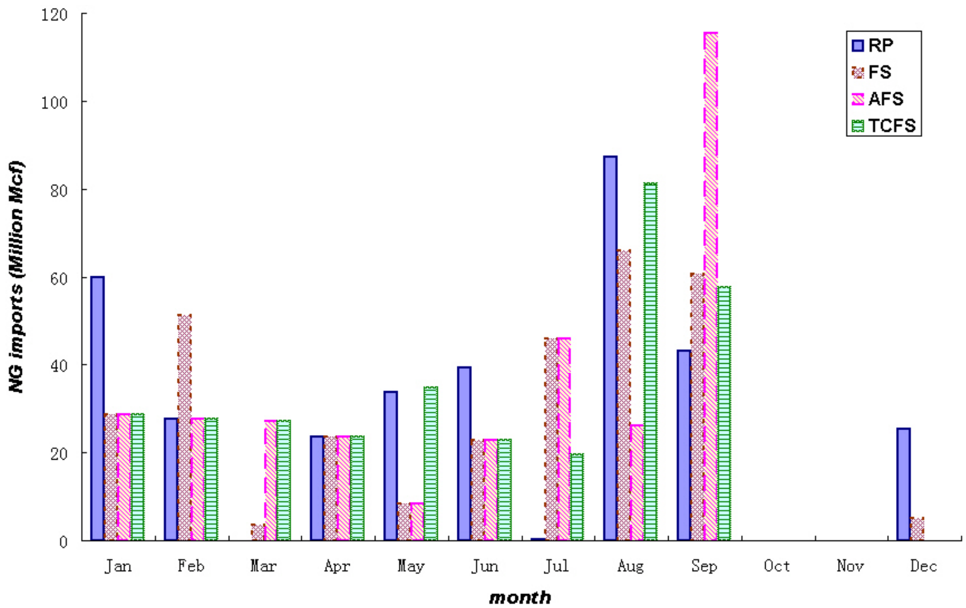

Figure 7.9 The NYNE model 2008 NG imports: RP vs. FS vs. AFS vs. TCFS

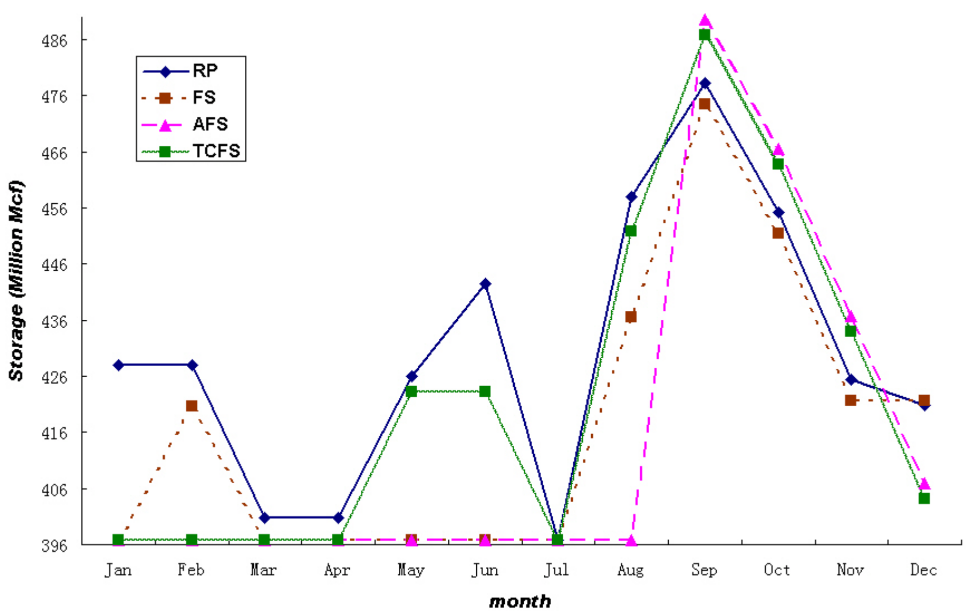

Figure 7.10 The NYNE model 2008 NG storage: RP vs. FS vs. AFS vs. TCFS

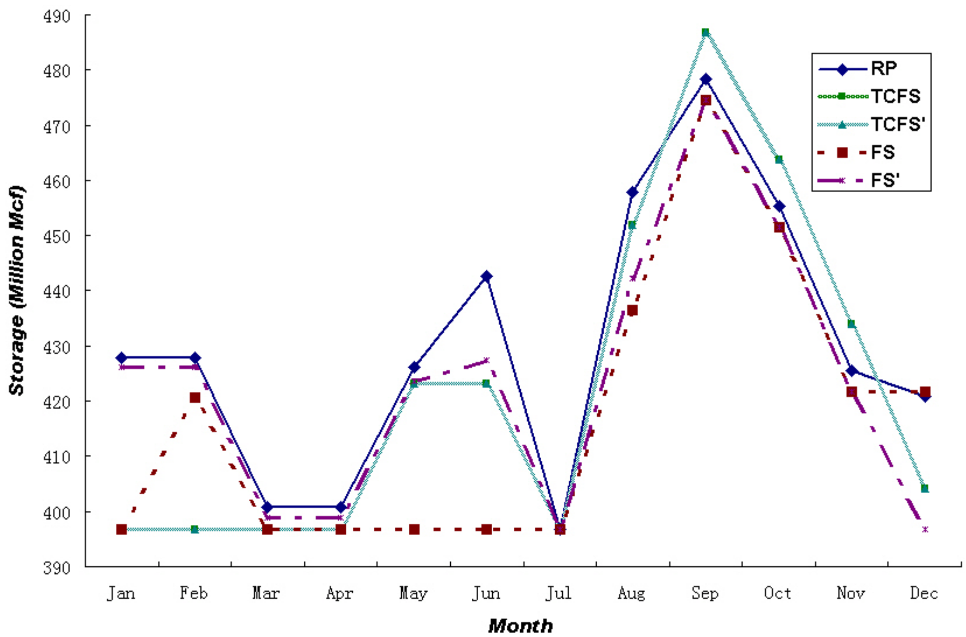

Figure 7.11 The NYNE model 2008 NG storage: variations of FS and TCFS

CHAPTER 8 CONCLUSIONS

Economic efficiency in the supply of electric energy is a decisive prerequisite for continued economic growth. To meet the demand for electrical energy, which increases by 4 to 7% per year in industrialized countries, considerable amounts of primary energy carriers such as coal, petroleum and natural gas must be provided for power generation. Power plants together with the production and transmission of fuels compose a complex network fraught with uncertainty. Due to the data availability and the complex interaction between subsystems, most energy models found in the literature have either a narrow geographic focus or a perspective limited to a single aspect of the whole system. Systems for the supply and transport of fuels and electric power therefore are investigated separately despite being highly interconnected. Especially, the modeling of uncertainty in medium-term models of national energy systems is still largely missing in the literature.

We used a stochastic program to model the uncertain fuel costs and better simulate the energy flows with a national scope. We examined how the inclusion of uncertainty affects the model's results; in particular, in historical case studies, whether this inclusion improves the model's accuracy in simulation. The computational efficiency issue is inherent with implementation of large-scale stochastic model because the size of a model usually grows dramatically when uncertainty is considered. We devised scenario reduction algorithms to address the tremendous computation task encountered in the large-scale optimization problem.

8.1 A large-scale stochastic program for energy flows under fuel cost uncertainty

To explore and simulate decision making in the U.S. electric energy system under uncertainty, uncertain fuel costs were included in a model of the bulk energy transportation system, which is composed by coal, natural gas and electricity subsystems and validated with actual data. We modeled the uncertain elements as discretely distributed random variables and used a two-stage approach to solve the stochastic problem. A small electricity network example illustrated the two-stage method and the difference between the flows in the stochastic model and those in the deterministic model.

To simulate decision-making in case studies of two separate years, the two-stage approach was applied in a rolling procedure to solve the multi-period problem, in which the fuel costs are revealed and forecasts are updated period by period. The scenarios of the natural gas costs were derived from a trusted source of forecasts. Due to the large number of scenarios, the implemented case studies have multi-million variables and constraints, which makes it non-trivial to solve the problem on a regular PC with limited memory. The optimal solution to the largest deterministic equivalent is obtained via Benders decomposition, the convergence time for which is problem dependent and highly correlated to the initial point. In order to ease the computational load and accelerate calculation speed, three approximation methods are employed and compared. Both temporal aggregation and structural scenario reduction accomplished considerable problem size reduction and maintained accuracy of the conclusions drawn from the RP approximations. Without making use of the model features, the general scenario reduction method costs greater computational effort.

The model was first tested with 2002 data. Compared to the recourse problem solution, the expected value solution that is obtained from the deterministic model with expected future fuel costs was closer to the optimal solution with perfect information. However, the recourse problem solution, which includes more natural gas consumption, less inter-regional electricity trade and higher natural gas storage levels, was similar to what actually happened in year 2002. Observations of more balanced use of fuels and procurement from additional supply areas in the recourse solution were consistent with theory that predicts greater diversification in solutions to stochastic optimization models as a hedge against uncertainty. Moreover, the results from a version of the model with load decomposition indicates that the difference between the solution to the deterministic (true value) solution with perfect information and the actual flows should not be attributed mainly to temporal aggregation of electric load. The solution from the stochastic model is stable under increased uncertainty. Therefore, the guesswork of setting the probabilities and the width of confidence intervals has a small effect on the outcome of the stochastic solution. Then the results from 2006 data confirm the impacts of the introduction of uncertain fuel costs on the flows of energy supply, trade and storage within the model.

8.2 Scenario reduction algorithms for computation efficiency

While the first part of the dissertation is an application of stochastic programming model in the U.S. power system with the methodology and illustration composed into a journal paper [71], the second part is a more theoretical study on computational efficiency problem in large-scale SP models that is raised in the first part. In Chapter 6, we investigated several methods to ease the computational load and accelerate calculation speed for the large-scale problem. Temporal aggregation and structural scenario reduction are two strategies that making use of the model features and both accomplished considerable problem size reduction while maintaining accuracy of the conclusions. The general scenario reduction method obtains more accurate approximations but costs greater computational efforts. There is apparently a trade-off between accuracy and computation time.

In Chapter 7, we discussed the three major restrictions of the general scenario reduction methods, based on which two heuristic algorithms are proposed. The accelerated forward selection heuristic is an excellent approximation of the original forward selection method and greatly exceeds FS in calculation efficiency. While most of the literature dealing the large scale problems measures the performance in terms of the total expected cost/benefit, we focus on the first-stage decisions instead. In the proposed TCFS method, the impact of a scenario on key first-stage variables is employed as the criterion for dividing scenarios into smaller groups. The clustering process reduces the computational burden and the incorporation of problem information improves the accuracy of the final approximation of the first-stage decisions. Moreover, TCFS method's performance is more stable than FS when the uncertain parameters involve different scales of quantities. We applied FS, AFS and TCFS to a three-period uncertain costs and demands problem. Both AFS and TCFS outperform FS in computation speed. TCFS obtains the most accurate solution. The two proposed algorithms are also applied to select scenarios for a regional (New York and New England) realization of the stochastic program in Chapter 3. Given its high efficiency in selecting scenarios, within the same quantity of time, more scenarios can be selected by TCFS than by FS. TCFS allows one representative scenario for every group of scenarios that have identical key first-stage decisions. The solution obtained with scenarios selected by TCFS is evidently closer to the exact solution than the approximation obtained by using either FS or AFS. We have shown in Chapter 6 that for both 2002 and 2006 cases, FS is accurate in estimating the total fuel usages. In Chapter 7, AFS and TCFS are applied to solve for approximations of the two large-scale problems. AFS reduces computation time from 1 day to less than one hour. TCFS obtains a more accurate estimation in terms of allocating NG usage to foreign and domestic supply areas.

APPENDIX A ILLUSTRATION OF THE ROLLING PROCEDURE WITH UPDATING FORECASTS

We use a three-period problem to illustrate the rolling procedure with updating forecasts. There is a single arc $(1,1)$ with uncertain costs $c_{11}(t), t = 2,3$ (the subscript 11 is supressed in the following text). Figure A.1 depicts the scenario tree constructed in the first period. The actual value of c for the first period is known as $c^a(1)$. Each uncertain cost has two possible values. Under the assumption of independence, they form four scenarios $S(1) = \{11, 12, 21, 22\}$ such that $c_1^{11}(2) = c_1^{12}(2) = c_1^1(2)$, $c_1^{21}(2) = c_1^{22}(2) = c_1^2(2)$, $c_1^{11}(3) = c_1^{21}(3) = c_1^1(3)$, and $c_1^{12}(3) = c_1^{22}(3) = c_1^2(3)$. The subscript 1 refers to the scenarios that are constructed in the first period. We solve the two-stage stochastic programming problem (A.1) and keep the first-stage decision $z(1)$, which is part of the RP solution. The alternative is to solve for the EV solution $\bar{x}(1)$ from the deterministic problem (A.2).

$$\begin{aligned}
\min \quad & c^a(1)z(1) + p_1^1(2)p_1^1(3)[c_1^1(2)y_1^{11}(2) + c_1^1(3)y_1^{11}(3)] + p_1^1(2)p_1^2(3)[c_1^1(2)y_1^{12}(2) + c_1^2(3)y_1^{12}(3)] \\
& +p_1^2(2)p_1^1(3)[c_1^2(2)y_1^{21}(2) + c(3)_1^1 y_1^{21}(3)] + p_1^2(2)p_1^2(3)[c_1^2(2)y_1^{22}(2) + c_1^2(3)y_1^{22}(3)] \\
\text{s.t.} \quad & a(1)z(1) + a(2)y_1^{11}(2) + a(3)y_1^{11}(3) = b \\
& a(1)z(1) + a(2)y_1^{12}(2) + a(3)y_1^{12}(3) = b \\
& a(1)z(1) + a(2)y_1^{21}(2) + a(3)y_1^{21}(3) = b \\
& a(1)z(1) + a(2)y_1^{22}(2) + a(3)y_1^{22}(3) = b
\end{aligned} \tag{A.1}$$

$$\min \quad c^a(1)\bar{x}(1) + \bar{c}_1(2)\bar{y}_1(2) + \bar{c}_1(3)\bar{y}_1(3)$$

$$\text{s.t.} \quad a(1)\bar{x}(1) + a(2)\bar{y}_1(2) + a(3)\bar{y}_1(3) = b \tag{A.2}$$

The acutal value of $c(2)$ is revealed at the beginning of the second period. Not only $c^a(2)$ falls out side of the predictions but also the forecasts for $c(3)$ update. The new set of scenarios is $S(2) = \{1, 2\}$. The renewed two-stage problem (A.3) is solved and $z(2)$ is obtained. Similarly to the previous period, we get the EV decision $\bar{x}(2)$ from problem (A.4). We are able to observe the true value of $c(3)$, which again does not coincide with any of the predictions, in the last period. Without any uncertainty, we solve the two deterministic problems (A.5) and (A.6), and complete RP as $\{z(1), z(2), z(3)\}$ and EV as $\{\bar{x}(1), \bar{x}(2), \bar{x}(3)\}$. For a perfect foresight benchmark, we find the optimal decision $TV = \{x(1), x(2), x(3)\}$ by solving the problem (A.7). Finally, the costs of all the solutions are evaluated as (12)–(14).

$$\begin{aligned} \min \quad & c^a(2)z(2) + p_2^1(3)c_2^1(3)y_2^1(3) + p_2^2(3)c_2^2(3)y_2^2(3) \\ \text{s.t.} \quad & a(2)z(2) + a(3)y_2^1(3) = b - a(1)z(1) \\ & a(2)z(2) + a(3)y_2^2(3) = b - a(1)z(1) \end{aligned} \tag{A.3}$$

$$\begin{aligned} \min \quad & c^a(2)\bar{x}(2) + \bar{c}_2(3)\bar{y}_2(3) \\ \text{s.t.} \quad & a(2)\bar{x}(2) + a(3)\bar{y}_2(3) = b - a(1)\bar{x}(1) \end{aligned} \tag{A.4}$$

$$\begin{aligned} \min \quad & c^a(3)z(3) \\ \text{s.t.} \quad & a(3)z(3) = b - a(1)z(1) - a(2)z(2) \end{aligned} \tag{A.5}$$

$$\begin{aligned} \min \quad & c^a(3)\bar{x}(3) \\ \text{s.t.} \quad & a(3)\bar{x}(3) = b - a(1)\bar{x}(1) - a(2)\bar{x}(2) \end{aligned} \tag{A.6}$$

$$\begin{aligned} \min \quad & c^a(1)x(1) + c^a(2)x(2) + c^a(3)x(3) \\ \text{s.t.} \quad & a(1)x(1) + a(2)x(2) + a(3)x(3) = b \end{aligned} \tag{A.7}$$

$$c^a(1)z(1) + c^a(2)z(2) + c^a(3)z(3) \tag{A.8}$$

$$c^a(1)\bar{x}(1) + c^a(2)\bar{x}(2) + c^a(3)\bar{x}(3) \tag{A.9}$$

$$c^a(1)x(1) + c^a(2)x(2) + c^a(3)x(3) \tag{A.10}$$

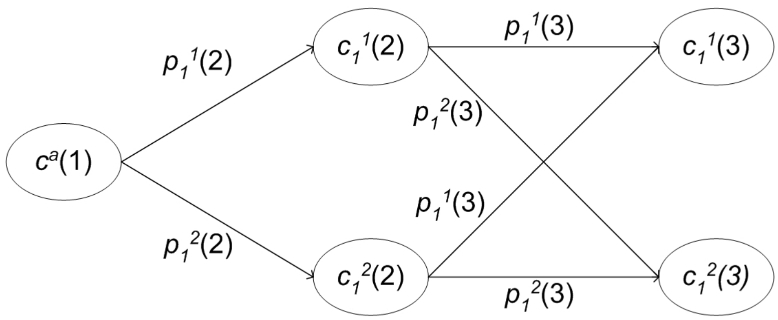

Figure A.1 Scenario tree: the first period

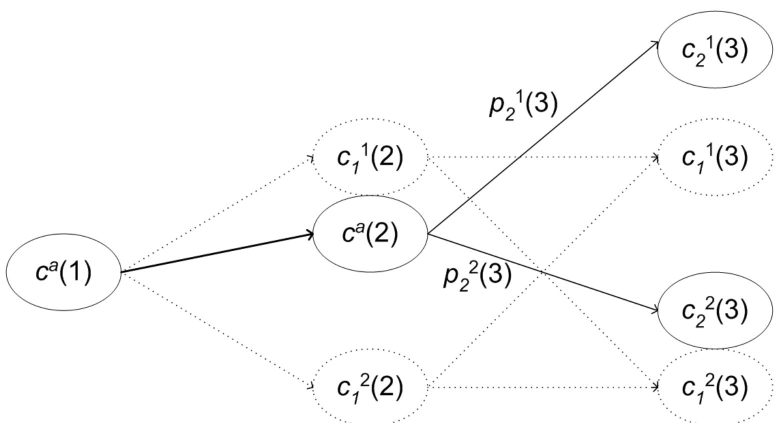

Figure A.2 Scenario tree: the second period

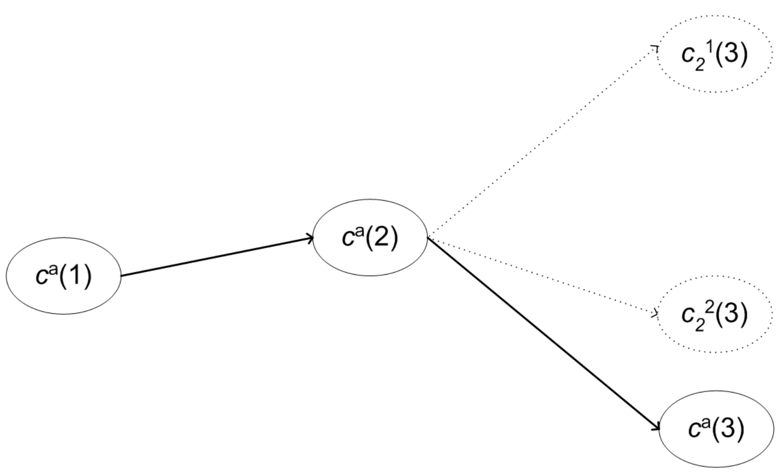

Figure A.3 Scenario tree: the third period

APPENDIX B THE MULTIPERIOD NETWORK FLOW MODEL OF BULK ENERGY TRANSPORTATION SYSTEM IN U.S.

B.1 List of nodes

Table B.1 lists all the nodes in the network flow model with 2002 data.

B.2 List of arcs

Table B.2 lists all the arcs in the network flow model with 2002 data.

B.3 Figures of the network

Table B.1 The list of nodes of the network flow model with 2002 data

Subsystem	Node	Representation	RHS	Distinct node #	period #	Total node #
Coal	XC	Dummy coal production node	N/A	1	1	1
Coal	CP	Aggregated coal production sites	0	11	1	11
Coal	CT	Virtual coal transmisstion centers	0	70	1	70
NG	XCA	Dummy NG production node for Canada	N/A	1	12	12
NG	GA	Aggregated NG production sites in Canada	0	4	12	48
NG	XG	Dummy domestic NG production nodes	0	14	12	168
NG	GT	Virtual NG transmission centers	0	6	12	72
NG	GS	Aggregated NG storage facilities	0	6	12	72
Elec	ED	Aggregated coal-powered generation sites, dry FGE	0	10	1	10
Elec	EN	Aggregated coal-powered generation sites, no FGD	0	17	1	17
Elec	EW	Aggregated coal-powered generation sites, wet FGD	0	17	1	17
Elec	EM	Aggregated NG-powered generation sites, CC	0	17	12	204
Elec	EB	Aggregated NG-powered generation sites, CT	0	17	12	204
Elec	EG	Aggregated coal-powered generation sites, GT	0	15	12	180
Elec	EL	Aggregated power transmission centers	d*	17	12	204

*Aggregated electricity demand in the region

FGD = flue gas desulfurization; CC = combined cycle; CT = combustion turbine; GS = gas steam.

Table B.2 The list of arcs of the network flow model with 2002 data

Arc Name	From	To	Representation	Cost	Coefficient	LB	UB	Distinct arc #	period #	Total arc #
XCCP	XC	CP	Aggregated domestic coal production	MP	1	0	PC	11	1	11
CPCT	CP	CT	Coal suply from mines to virtual transmission center	**	1	0	UB	70	1	70
CTED	CT	ED	Coal transffered from coal centers to ED coal plants	0	Coal h/p	0	RGC	37	1	37
CTEN	CT	EN	Coal transffered from coal centers to EN coal plants	0	Coal h/p	0	RGC	70	1	70
CTEW	CT	EW	Coal transffered from coal centers to EW coal plants	0	Coal h/p	0	RGC	70	1	70
XCAGA	XCA	GA	Aggregated NG wells in Canada	WP	1	0	PC	4	12	48
GAGT	GA	GT	NG supply from Canada	***	TL	0	PPC	4	12	48
XGGT	XG	GT	Aggregated domestic NG wells	WP	EL	0	PC	14	2	168
GTGT	GT	GT	NG transmissed among regions	***	TL	0	PPC	15	12	180
GSGT	GS	GT	NG supply from storage	SF	1	0	WC	6	12	72
GTGS	GT	GS	Store extra NG in the facilities	0	1	0	EC	6	12	72
step	GS	GS	NG storage from one period to the next	0	1	CG	SC	6	11	66
GTEM	GT	EM	NG transffered from NG centers to EM NG plants	0	NG h/p	0	UB	22	12	264
GTEB	GT	EB	NG transffered from NG centers to EB NG plants	0	NG h/p	0	UB	22	12	264
GTEG	GT	EG	NG transffered from NG centers to EG NG plants	0	NG h/p	0	UB	19	12	228
EDEL	ED	EL	Power generated by ED coal plant in each region	0	1	0	RGC	10	12	120
ENEL	EN	EL	Power generated by EN coal plant in each region	0	1	0	RGC	17	12	204
EWEL	EW	EL	Power generated by EW coal plant in each region	0	1	0	RGC	17	12	204
EMEL	EM	EL	Power generated by EM NG plant in each region	0	1	0	RGC	17	12	204
EBEL	EB	EL	Power generated by EB NG plant in each region	0	1	0	RGC	17	12	204
EGEL	EG	EL	Power generated by EG NG plant in each region	0	1	0	RGC	15	12	180
ELEL	EL	EL	Power transmissed among regions	TC	TL	0	ITC	58	12	696

** Difference between delivery price and minemouth price

*** Difference between delivery price and wellhead price

MP = Minemouth price; WP = wellhead price; SF = Fixed storage fee; TC = Fixed transferring cost.

h/p = heat value/plant heat rate; TL = Transmission loss; EL = Extraction loss.

CG = Cushion gas.

PC = Production capacity; UB = Unbounded; PPC = Pipeline capacity; WC = Withdraw capacity.

IC = Injection capacity; SC = Storage Capacity; RGC = Regional generation capacity; ITC = Interregional transmission capacity.

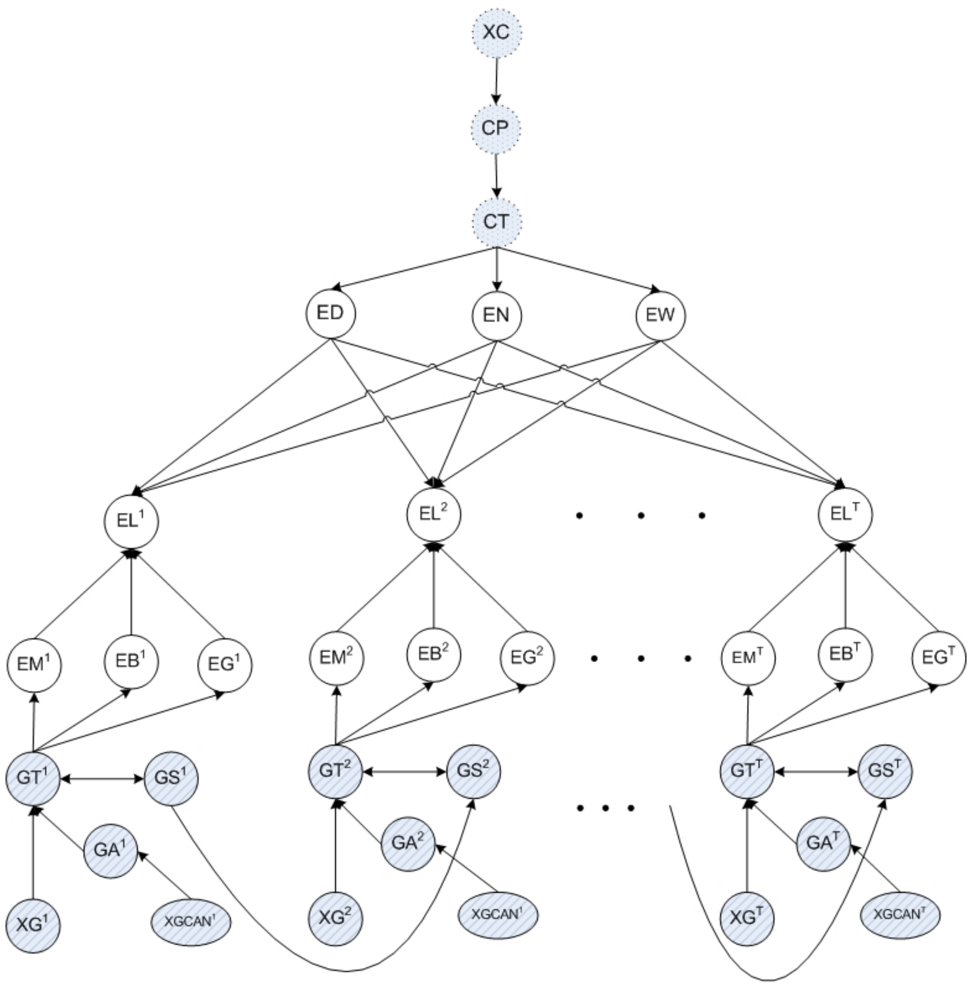

Figure B.1 The integrated power transportation system

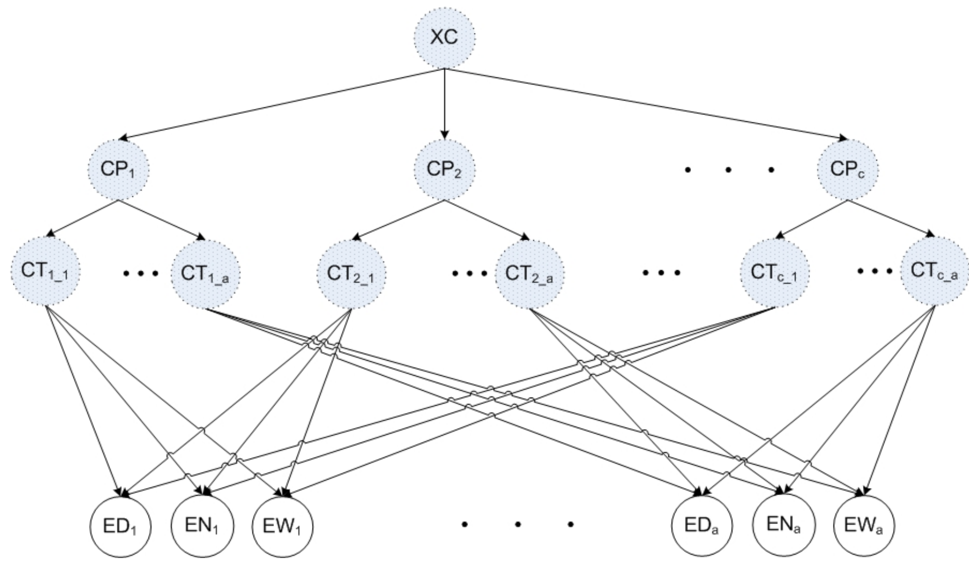

Figure B.2 The coal subsystem

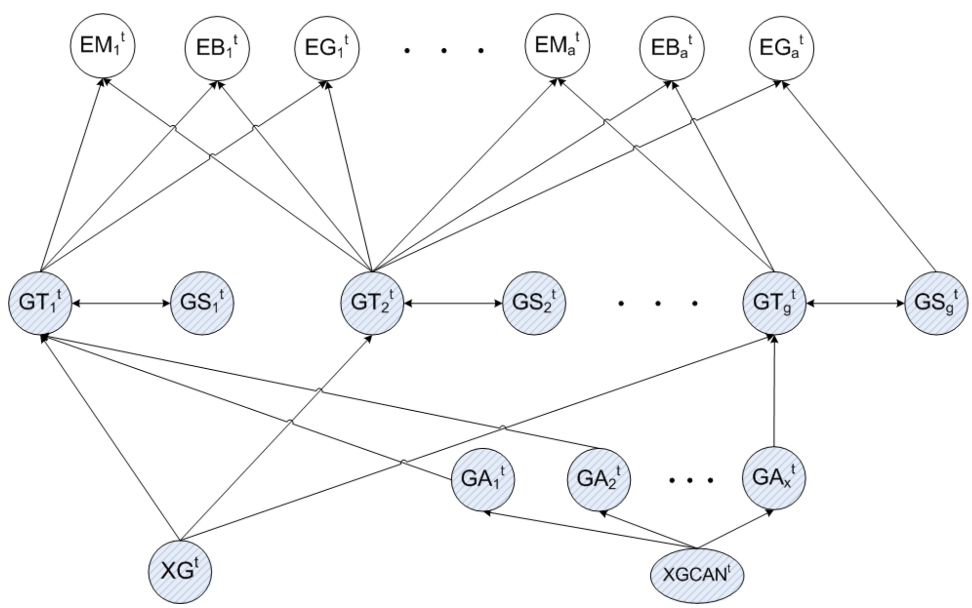

Figure B.3 The natural gas subsystem

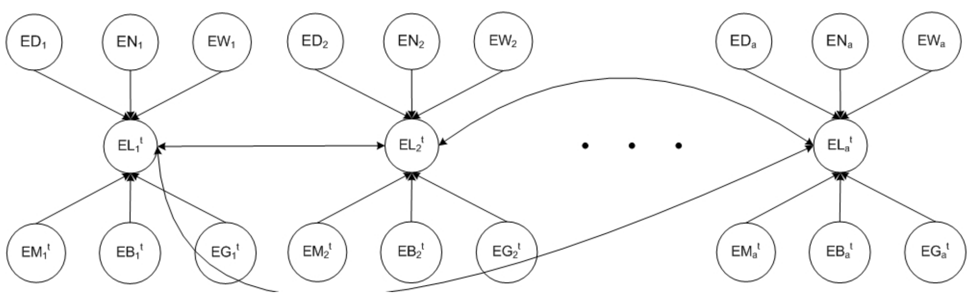

Figure B.4 The electricity subsystem

Bibliography

[1] Anderson, E. and Philpott, A. (2002). *Decision Making under Uncertainty: Energy and Power*, volume 128 of *IMA volumes on Mathematics and its Applications*, chapter On supply function bidding in electricity markets, pages 115–134. Springer, New York.

[2] Argonne (2009). Energy and power evaluation program (enpep-balance). Web.

[3] Beale, E. M. L. (1955). On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society*, B(17):173–184.

[4] Beltratti, A., Consiglio, A., and Zenios, S. (1999). Scenario modeling for the management of international bond portfolios. *Annals of Operations Research*, 85(1):227–247.

[5] Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.

[6] Benders, J. F. (1962). Partitioning procedures for solving mixed integer variables programming problems. *Numerische Mathematik*, 4:238–252.

[7] Bertsimas, D. and Sim, M. (2003). Robust discrete optimization and network flows. *Mathematical Programming*, 98(1–3):49–71.

[8] Birge, J. J. M. (1996). *Mathematical programming for industrial engineerings*, chapter Stochastic programming, pages 543–574. Dekker, New York.

[9] Bolinger, M., Wiser, R., and Golove, W. (2006). Accounting for fuel price risk when comparing renewable to gas-fired generation: the role of forward natural gas prices. *Energy Policy*, 34:706–720.

[10] Botnen, O., A., J., Haugstad, A., Kroken, S., and Froystein, O. (1992). Modelling of hydropower scheduling in a national/international context. In Broch, E. and Lysne, D., editors, *Proceedings of the 2nc international Conference on Hydropower Development (Hydropower 92)*, pages 575–584. Lillehammer, Norway; Balkema, Rotterdam.

[11] Butler, J. C. and Dyer, J. S. (1999). Optimizing natural gas flows with linear programming and scenarios. *Decision Sciences*, 30(2):563–580.

[12] Canestrelli, E. and Giove., S. (1999). *Current topics in Quantitative Finance*, chapter Scenarios identification for financial time series, pages 25–36. Physica-Verlag.

[13] Carino, D., Myers, D., and Ziemba, W. (1998). Concepts, technical issues, and uses of the Russell-Yasuda Kasai financial planning model. *Operations Research*, 46(4):450–462.

[14] Casimir, R. (1990). The newsboy and the flower-girl. *Omega*, 18:395–398.

[15] Charnes, A. and Cooper, W. (1959). Chance-constrained programming. *Management Science*, 6(1):73–79.

[16] Dantzig, G. and Glynn, P. (1990). Parallel processors for planning under uncertainty. *Annals of Operations Research*, 22(1):1–21.

[17] Dantzig, G. and Infanger, G. (1993). Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research*, 45:59–76.

[18] Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science*, 1:197–206.

[19] Dembo, R. (1991). Scenario optimization. *Annals of Operations Research*, 30:63–80.

[20] Dempster, M. and Thompson, R. (1996). EVPI-based importance sampling solution procedures for multistage stochastic linear programs on parallel mimd architectures. *Annals of Operations Research*, 90:161–184.

[21] Dupačová, J., Gröwe-Kuska, N., and Römisch, W. (2003). Scenario reduction in stochastic programming. *Mathematical Programming*, 95(3):493–511.

[22] EIA (2002). Short-term energy outlook January 2002.

[23] EIA (2003a). The national energy modeling system: An overview.

[24] EIA (2003b). Short-term energy outlook January 2003.

[25] EIA (2006a). Annual energy review.

[26] EIA (2006b). Short-term energy outlook January 2006.

[27] EIA (2007). Short-term energy outlook January 2007.

[28] Fleten, S.-E., Wallace, S., and Ziemba, W. (2002). *Decision Making under Uncertainty: Energy and Power*, volume 128 of *IMA volumes on Mathematics and its Applications*, chapter Hedging electricity portfolios using stochastic programming, pages 71–94. Springer, New York.

[29] Fragniere, E. and Haurie, A. (1996). A stochastic programming model for energy/environment choices under uncertainty. *International Journal of Environment and Pollution*, 6(4–6):587–603.

[30] Gardner, D. and Rogers, J. (1999). Planning electric power systems under demand uncertainty with different technology lead times. *Management Science*, 45:1289–1306.

[31] Heitsch, H., Römisch, W., and Strugarek, C. (2007). Stability of multistage stochastic programs. *SIAM Journal of Optimization*, 17(2):511–525.

[32] Hindsberger, M. . (2003). *Interconnected hydro-thermal systems.* PhD dissertation, Technical University of Denmark.

[33] Hogan, W. W. (1975). Energy policy models for project independence. *Computers and Operations Research*, 2:251–271.

[34] Høyland, K. and Wallace, S. W. (2001). Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307.

[35] Infanger, G. (1992). Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. *Annals of Operations Research*, 39(1):65–95.

[36] Jacobs, J., Freeman, G., Grygier, J., Morton, D., Schultz, G., Staschus, K., and Stedinger, J. (1995). Socrates: A system for scheduling hydroelectric generation under uncertainty. *Annals of Operations Research*, 59:99–132.

[37] Klassen, P. (1997). Discretized reality and spurious profits in stochastic programming models for asset/libility management. *European Journal of Operational Research*, 101:374–392.

[38] Klassen, P. (1998). Financial asset-pricing theory and stochastic programming models for asset/liability management: A synthesis. *Management Science*, 44:31–48.

[39] Koskosidis, Y. and Duarte, A. (1997). A scenario-based approach to active asset allocation. *Journal of Portfolio Management*, 23(2):74–85.

[40] Kouvelis, P. and Yu, G. (1997). *Robust Discrete Optimization and Its Application*. Kluwer Academic Publishers, Boston.

[41] Lavenberg, S. S. and Welch, P. D. (1981). A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science*, 27(3):322–335.

[MathWorks] MathWorks. Matlab r2010b documentation: Statistics toolbox.

[42] Mo, B., Gjelsvik, A., and Grundt, A. (2001). Integrated risk management of hydro power scheduling and contract management. *IEEE Transactions on Power Systems*, 16(2):216–221.

[43] Mulvey, J. and Vladimirou, H. (1998). *World Wide Asset and Liability Modeling*, chapter The towers Perrin global capital market scenario generation system, pages 286–312. Cambridge University Press.

[44] Mulvey, J. and Zenios, S. (1995). Capturing the correlations of fixed-income instruments. *Management Science*, 40:1329–1342.

[45] Mulvey, J. M. and Vladimirou, H. (1991). Solving multistage stochastic networks: an application of scenario aggregation. *Networks*, 21:619–643.

[46] Murphy, F. and Sen, S. (2002). *Decision Making under Uncertainty: Energy and Power*, volume 128 of *IMA volumes on Mathematics and its Applications*, chapter Qualitative implications of uncertainty in economic equilibrium models, pages 135–152. Springer, New York.

[47] Murphy, F., Sen, S., and Soyster, A. (1982). Electric utility capacity expansion planning with uncertain load forecasts. *IIE Transactions*, 14(1):52–59.

[48] Murphy, F. H., Conti, J. J., Shaw, S. H., and Sanders, R. (1988). Modeling and forecasting energy markets with the intermediate future forecasting system. *Operations Research*, 36(3):406–420.

[49] Neame, P., Philpott, A., and Pritchard, G. (1999). Offer stack optimization for price takers in electricity markets. In *Proceedings of the 34th Annual Conference of ORNZ*, pages 3–12, Hamilton, New Zealand. University of Waikato.

[50] Nowak, M. and Römisch, W. (2000). Stochastic lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. *Annals of Operations Research*, 100:251–272.

[51] Pereira, M. V. F. and Pinto, L. M. V. G. (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375.

[52] Pflug, G. C. (2001). Scenario tree generation for multiperiod financial optimization. *Mathematical Programming*, 89:251–271.

[53] Qiu, J. and Girgis, A. (1993). Optimization of power-system reliability level by stochasticprogramming. *Electric Power Systems Research*, 26(2):87–95.

[54] Quelhas, A., Gil, E., and McCalley, J. (2007a). A multiperiod generalized network flow

model of the U.S. integrated energy system: Part II–simulation results. *IEEE Transactions on Power Systems*, 22(2):837–844.

[55] Quelhas, A., Gil, E., McCalley, J., and Ryan, S. M. (2007b). A multiperiod generalized network flow model of the U.S. integrated energy system: Part I–model description. *IEEE Transactions on Power Systems*, 22(2):829–836.

[56] Quelhas, A. M. (2006). *Economics efficiencies of the flows from the primary resource suppliers to the electric load centers*. PhD dissertation, Iowa State University.

[57] Roh, J. H., Shahidehpour, M., and Wu, L. (2009). Market-based generation and transmission planning with uncertainties. *IEEE Transactions on Power Systems*, 24(3):1587–1598.

[58] Römisch, W. and Shapiro, A. (2003). *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, chapter Stability of Stochastic Programming Problem, pages 483–554. Elsevier.

[59] Ruszczyński, A. and Shapiro, A. (2003). *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, chapter Stochastic Programming Model, pages 1–64. Elsevier.

[60] Ryan, S. M., Downward, A., Philpott, A., and Zakeri, G. (2010). Welfare effects of expansions in equilibrium models of an electricity market with fuel network. *IEEE Transactions on Power Systems*, 25(3):1337–1349.

[61] Ryan, S. M. and Wang, Y. (Forthcoming). *Handbook of Networks in Power Systems*, chapter Costs and constraints of transporting and storing energy for dispatchable electricity generation: Implications for optimization models. Springer.

[62] Salas, J., III, G. T., and Bartolini, P. (1985). Approaches to multivariate modeling of water resources time series. *Water Resources Bulletin*, 21:683–708.

[63] Scott, T. and Read, E. (1996). Modelling hydro reservoir operation in a deregulated electricity sector. *International Transactions in Operations Research*, 3(3–4):209–221.

[64] Sherali, H., Soyster, A., Murphy, F., and Sen, S. (1982). Linear programming based analysis of marginal cost pricing in electric utility capacity expansion. *European Journal of Operational Research*, 11:349–360.

[65] Sherali, H., Soyster, A., Murphy, F., and Sen, S. (1984). Intertemporal allocation of capital costs in electric utility capacity expansion planning under uncertainty. *Management Science*, 30:1–19.

[66] Tuohy, A., Meibom, P., Denny, E., and OMalley, M. (2009). Unit commitment for systems with significant wind penetration. *IEEE Transactions on Power Systems*, 24(2):592–601.

[67] Van Slyke, R. M. and Wets, R. (1969). L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663.

[68] Wallace, S. W. and Fleten, S. E. (2003). *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, chapter Stochastic Programming Models in Energy, pages 637–677. Elsevier.

[69] Wang, J., Shahidehpour, M., and Li, Z. (2008). Security-constrained unit commitment with volatile wind power generation. *IEEE Transactions on Power Systems*, 23(3):1319–1327.

[70] Wang, Y. and Ryan, S. M. (2009). Comparison of efficient methods for solving a large-scale multistage stochastic program. In *Proceedings of the 2009 Industrial Engineering Research Conference*. Institute of Industrial Engineers.

[71] Wang, Y. and Ryan, S. M. (2010). Effects of uncertain fuel costs on fossil fuel and electric energy flows in the U.S. *Energy Systems*, 1:209–243.