

The Pennsylvania State University

The Graduate School

Harold and Inge Marcus

Department of Industrial and Manufacturing Engineering

**MULTI-STAGE STOCHASTIC PROGRAMMING ON
SHALE GAS INFRASTRUCTURE AND PRODUCTION PLANNING**

A Thesis in

Industrial Engineering and Operations Research

by

Xiaohuang Wu

© 2012 Xiaohuang Wu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2012

The thesis of Xiaohuang Wu was reviewed and approved* by the following:

Tao Yao
Associate Professor of Industrial and Manufacturing Engineering
Thesis Advisor

Jack C. Hayya
Professor Emeritus of Management Science and Information Systems

Paul Griffin
Professor of Industrial and Manufacturing Engineering
Peter and Angela Dal Pezzo Department Head Chair

*Signatures are on file in the Graduate School

ABSTRACT

The rich resource of Marcellus gas has recently boosted up the interest of people as the drilling technology advances. Apart from the difficulties in exploration and drilling for natural gas, problems of building and developing network that delivers the gas to customers remain unsolved for shale gas companies and government. In the paper, we investigate an in-land shale gas infrastructure and production planning problem. Multiple gas fields with uncertain reserves are considered. Platforms are connected by pipelines, which will eventually transport the gas to a central pipeline that delivers gas to merchants and customers. The goal is to obtain the maximum expected net present value of the project within a given time horizon. As the revelation of field reserves will affect the decision maker's action, we develop a stochastic model with endogenous uncertainties. Modern stochastic programming technique is applied. Specifically, conservative approximations are obtained under the assumption of piecewise constant binary and linear real-valued decision rules. The decision rule approximation successfully solves a small sized example with continuously distributed uncertainty parameters. Near-optimal results are obtained efficiently and can be improved by increasing partition subsets.

TABLE OF CONTENTS

List of Figures	v
List of Tables	vi
Acknowledgements	vii
Chapter 1 Introduction	1
Chapter 2 Literature Review	4
Chapter 3 Problem Statement	7
Chapter 4 One-stage Infrastructure and Production Planning	12
Model	12
Approximation	13
Numerical Example	15
Chapter 5 Multi-stage Infrastructure and Production Planning	21
Model	21
Approximation	22
Numerical Example	26
Chapter 6 Conclusions and Future Work	40
List of References	41

List of Figures

Figure 3-1. Shale gas network.	7
-------------------------------------	---

List of Tables

Table 3-1. Notation.	8
Table 4-1. Input parameters for one-stage.	16
Table 5-1. Input parameters for multi-stage.	28
Table 5-2. Partition of uncertainty set.....	31
Table 5-3. Optimal construction.	34
Table 5-4. Optimal gas extraction.....	35
Table 5-5. Optimal capacity expansion.	36
Table 5-6. Optimal gas flow.	37

Acknowledgements

First, I want to thank my thesis advisor, Dr. Yao. It is his guidance, insights and patience that make it possible for me to rise in the academic world, one full of joy and surprise that I have never experienced before. I also want to thank my thesis reader Dr. Hayya and department chair Dr. Griffin for spending their time reading my thesis and providing me with valuable feedback. Finally, I want to thank my family and my friends. Without their persistent support, I could not have made it this far. They are and will always be the most important part of my life.

Chapter 1

Introduction

The Marcellus shale, a layer of shale rock beneath the rolling hills and mountains of Pennsylvania, is the largest unconventional natural gas reserve in the world. This well-known geological formation that contains significant amounts of natural gas was never considered worthwhile until recent technology advances. Though reserve estimates are considered uncertain at this point, most of the completed Marcellus well revealed abundant recoverable reserves. Together with the emphasis on consuming green energy due to environmental issues, natural gas, which has considerably lower carbon content than petroleum and coal, will no doubt grasp its own share of the market.

Marcellus shale gas play becomes more and more attractive these days as the drilling technology advances. Reports show that the activity in the Marcellus will continue to expand and natural gas production from Marcellus could rise to almost 4 billion cubic feet BCF per day by 2020 (Considine and Watson 2009). However, among some key factors affecting development is infrastructure and production planning. Though it may seem hard to explore and drill for natural gas, the real work happens to be the development of a network consisting of thousands of miles of gathering lines and pipelines to carry the gas to consumers (Considine et al. 2010). Besides, building natural gas processing facilities takes considerable time and incurs significant costs. To sum up, developing transportation and production processing networks takes money and time. But the high potential environment and investment impact stimulates the research.

One of the most major concerns is that the decision maker is exposed to a great deal of uncertainty. Though the location of gas fields could be identified, the amount of gas in these reserves remains uncertain until the platform is built. Therefore, it is crucial to take uncertainties into account when formulating the model.

A typical shale gas infrastructure and production problem has multiple potential reserves within a region to build well platforms on. The extracted natural gas needed to be purged and dried before sold to merchants and customers for use. Therefore, a production platform is usually built on site to process the gas before it can be transported to other areas. In the paper, we call the combo of well platform and production platform ‘a platform’ in general. After the gas is processed on site, it is transported via pipelines that connecting platforms together. All the gas produced is to be transported to a central pipeline that connects different regions together. The shale gas infrastructure and production planning problem requires the decision maker to select when and where to build platforms, to increase the capacity of platforms and to install pipelines. Besides, operation decisions are to be made to determine the production schedule for different gas fields.

For simplification, the planning horizon is discretized into time periods. The decision maker is to make decision at the beginning of each time period and all the decision take immediate effect. As the decisions regarding when and where to build platforms and pipelines are represented as binary variables, and other decisions including the capacity expansion and extraction schedule are real-valued, the formulation has the form of mixed binary integer programming. However, all the decision variables are dependent on the information of gas reserves. The problem thus becomes a stochastic

programming with endogenous uncertainties. In the following chapter, we will present both a one-stage and a multi-stage stochastic programming model for the shale gas infrastructure and production planning.

The paper is organized as follows:

In chapter 2, a literature review is presented regarding previous work on gas production planning. Research works on stochastic programming that are used to handle uncertainty are also summarized.

In chapter 3, the problem statement and assumptions are presented. The description of the model is also discussed.

In chapter 4 and 5, models are built for the one-stage and the multi-stage infrastructure and production planning respectively. An approximation of each one of the problems is presented followed by a numerical example with three gas fields and five links.

The last chapter consists of conclusions of the thesis together with some suggested future research directions.

Chapter 2

Literature Review

The economic impact of shale gas is studied by Considine et al. (2009, 2010). They use IMPLAN modeling system to estimate the job creation, value added, etc. However, the development of the shale gas network consisting of thousands of miles of gathering lines and pipelines to carry the gas to consumers remains a major concern (Considine et al. 2010).

Though the shale gas infrastructure and production planning problem is a relatively new topic, there have been intensive research studies regarding the oil and gas field infrastructures. Comprehensive studies on deterministic approaches can be dated back to 1998 (Ierapetritou, Floudas, Vasantharajan and Cullick, 1998; Iyer, Grossmann, Vasantharaja and Cullick, 1998; Grothey and McKinnon, 2000; Barnes, Linke and Kokossis, 2002; Kosmidis, Perkins and Pistikopoulos, 2002; Lin and Floudas, 2003; Ortiz-Gomez, Rico-Ramirez and Hernandez-Castro, 2002). Uncertainty has also been considered in some of the literatures (Haugen 1996; Meister, Clark and Shah 1996; Jonsbraten 1998; Goel and Grossmann 2004).

A dynamic stochastic programming model that incorporates with uncertainty in the size of oil fields is proposed by Haugen (1996). The author only considers the decisions made for scheduling of oil fields. Exploration and production decisions for one field under uncertainty in reserves and oil price are studied by Meister, Clark and Shah (1996). Jonsbraten (1998) uses the progressive hedging algorithm to make decisions for

an oil field with uncertainty in oil prices. The problem is formulated as a mixed integer linear programming. The author also studies the sequencing of oil wells under uncertainty in size of oil fields. Both of these two works only include one oil field.

Based on the dependence of the order of revelation of uncertainties on decision maker's action in stochastic programming, the uncertainties can be categorized as exogenous uncertainties and endogenous uncertainties (Jonsbraten 1998). Jonsbraten (1998) investigates decision problems in which actions can affect both the distribution of the uncertainties and the timing of revelation. Besides, there are research studies focusing on problems based on the assumption that uncertain parameters follow a discrete distribution, which could be solved using a finite scenario tree. Goel and Grossmann (2004) propose a model for off-shore gas field development problem with discretely distributed endogenous uncertainties and reformulate it as a mixed binary program.

Problems with continuously distributed random parameters, otherwise, need to be discretized before any of the above techniques can be applied. One alternate solution could be Monte Carlo sampling, but such technique is not considered computationally efficient (Dyer and Stougie 2006, Shapiro and Nemirovski 2005). While discretization may be able to get accurate approximations, it can increase computing cost dramatically when applied to medium or large-sized problems. Vayanos and Kuhn (2011) proposed a methodology for solving dynamic problems with endogenous uncertainties. They suggest approximating the adaptive measurement decisions by piecewise constant functions and the adaptive real-valued decisions by piecewise linear functions of the uncertainties.

The contribution of this paper is as follows:

First, in this paper, we seek to propose models suitable for in-land shale gas infrastructure and production planning, which has not been studied before. A one-stage model and a multi-stage model with endogenous uncertainties are presented respectively. The assumptions and restrictions of the model are discussed.

Second, most of the past research works in oil field development planning under uncertainty deal with discretely distributed random parameters or a single oil field. We explore the problem in a broader context as multiple oil fields with continuously distributed uncertain reserves are presented and the project can last for multiple time horizons.

Finally, we provide a stochastic programming technique to solve the multi-region and multi-stage dynamic decision making problem efficiently by partition of uncertainty set. Near-optimal conservative approximations are obtained under the assumption of piecewise constant binary and linear real-valued decision rules.

Chapter 3

Problem Statement

A gas company has several region for gas extraction, and for each region, the gas company has located several gas fields with unknown reserves. As for the case with inland shale gas production, the production platform is always built next to a gas field that the company decides to exploit. The gas will be processed on-site at each production platform and then transported to a main pipeline. Suppose the main pipeline is built and known and the company is assumed to be aware of the exact locations of the individual gas fields. The company needs to plan the gas production process and build the pipeline to transport the gas to the main pipeline.

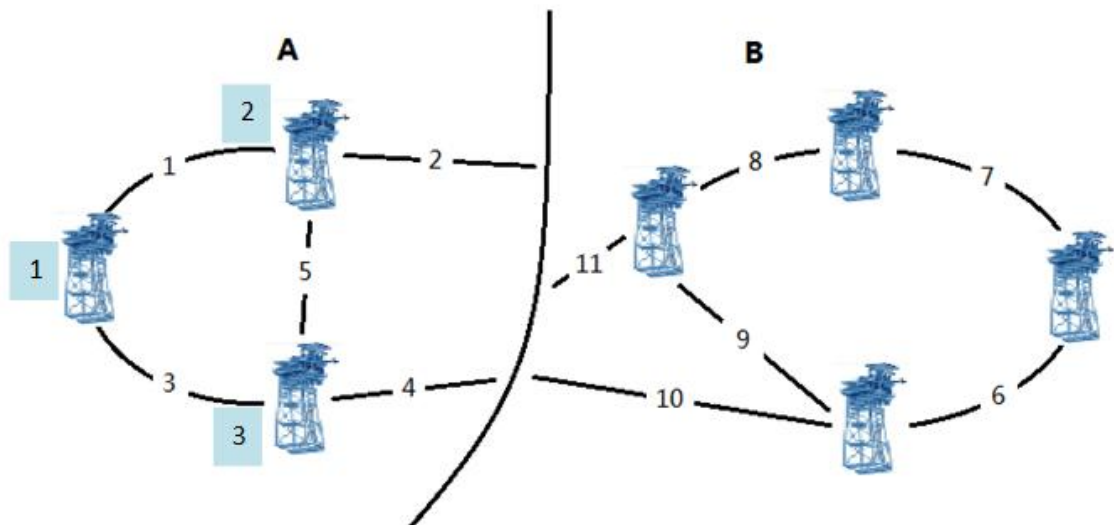


Figure 3-1. Shale gas network.

The model's formulation requires the notation in table 3-1.

Table 3-1. Notation.

Γ	Time horizon, in years
P	The set of candidate production platforms.
L	The set of possible pipelines between production platforms
o	Main pipeline
$L^+(p)$	The set of all ingoing pipeline to production platform p
$L^-(p)$	The set of all outgoing pipeline from production platform p
r_p	Maximum production rate at production platform p
d_t	Discount factor at year t
c_g	Unit price for gas
c_c^p	Unit capacity expansion cost for production platform p
c_e^p	Unit gas extraction cost for platform p
c_i^l	Cost for building pipeline l
c_i^p	Cost for building platform p
ξ^p	Gas field size. Random variable.
x_t^p	Binary variable. =1 if production platform p exists in year t
x_t^l	Binary variable. =1 if pipeline l exists in year t
$y_{e,t}^p$	The amount of gas extraction for production platform p in year t

$y_{f,t}^l$	The amount of gas flow through pipeline l in year t
$y_{c,t}^p$	The capacity of production platform P that is increased at the beginning of year t

Assume the company has a project horizon of T years, the time horizon is discretized into T segments, with each one equals to a year. The company makes decisions on platform expansion, gas extraction and platform and pipeline construction at the beginning of each year.

To extract gas from a certain gas reserve, platform has to be installed at the corresponding field. Installed pipelines and platforms are not to be salvaged. Capacity expansion and gas extraction schedules may differ from year to year. All decisions take immediate effect. Once platform is built at P , the size of the gas field ξ^p will be revealed. The total amount of gas extracted should not exceed the reserve of the gas field ξ^p . Once information is revealed, it is assumed not to be forgotten. Annual gas production is also limited by a production rate which may vary field by field, depending on the platform.

The uncertainty is characterized by the reserve of each gas field, which follows a continuous distribution. The company is to make investment and operational decisions each year based on the information it has up to each corresponding year. As random parameters are continuously distributed and decisions are dependent of the uncertainty, a dynamic tree model will not be suitable for the problem. As the uncertainties are unfold over the entire time horizon T based on the investment decisions and operational

decisions are dynamic and spread over the entire time horizon, recourse decisions should be considered.

The problem could be stated in detail as below.

The company's goal is to maximize the expected net present value of the project.

The profit for a single year equals to the sales of gas less the construction cost of pipelines and platforms less the capacity expansion cost at each platform installed less the gas extraction cost. Note that the profit at a later year will be discounted by a factor d_t before added to the objective function.

$$z = \sum_{t \in \Gamma} d_t E \left[c_g \sum_{l \in L^+(o)} y_{f,t}^l(\xi) - \sum_{l \in L} c_i^l (x_t^l(\xi) - x_{t-1}^l(\xi)) - \sum_{p \in P} c_i^p (x_t^p(\xi) - x_{t-1}^p(\xi)) - c_c^p y_{c,t}^p(\xi) - c_e^p y_{e,t}^p(\xi) \right].$$

1. The total amount of gas extracted over the entire time horizon from a single gas field should not exceed the gas field size. Thus,

$$\sum_{t \in \Gamma} y_{e,t}^p(\xi) \leq \xi^p \quad \forall p \in P.$$

2. Gas production is limited by a maximum production rate at a particular production platform. Hence,

$$0 \leq y_{e,t}^p(\xi) \leq r_p \quad \forall p \in P.$$

3. The network subjects to the flow conservation constraint. So,

$$y_{e,t}^p(\xi) + \sum_{l \in L^-(p)} y_{f,t}^l(\xi) \geq \sum_{l \in L^+(p)} y_{f,t}^l(\xi) \quad \forall p \in P.$$

4. Gas flow from a particular production platform should not exceed its capacity.

Hence,

$$\sum_{l \in L^-(p)} y_{f,t}^l(\xi) \leq \sum_{\tau=1}^t y_{c,\tau}^p(\xi) \quad \forall p \in P, t \in \Gamma.$$

5. No gas flow from pipeline l if the pipeline has not been built. Thus,

$$0 \leq y_{f,t}^l(\xi) \leq Mx_t^l(\xi) \quad \forall l \in L, t \in \Gamma.$$

6. No expansion can be constructed if production platform P has not been built.

It follows that

$$0 \leq y_{c,t}^p(\xi) \leq Mx_t^p(\xi) \quad \forall p \in P, t \in \Gamma.$$

7. Existing pipelines and production platforms are not to disappear in the network.

Thus,

$$x_{t-1}^p(\xi) \leq x_t^p(\xi) \quad \forall p \in P, t \in \Gamma,$$

$$x_{t-1}^l(\xi) \leq x_t^l(\xi) \quad \forall l \in L, t \in \Gamma.$$

Chapter 4

One-stage Infrastructure and Production Planning

Model

Notation. In this chapter, uncertainty is modeled by a probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P)$ that consists of the sample space \mathbb{R}^k and the Borel σ -algebra $\mathcal{B}(\mathbb{R}^k)$, which is the set of events that are assigned probabilities by the probability measure P . Let $M_{k,n}$ denote the space of all measurable functions from \mathbb{R}^k to \mathbb{R}^n . Let $E(\cdot)$ denote the expectation operator with respect to P and $x \bullet y$ denote the Hadamard product of two vectors $x, y \in \mathbb{R}^n$.

This section discusses the one-stage infrastructure and production planning problem with endogenous uncertainty. The decision maker first selects some gas fields, i.e. some elements within ξ to observe. The construction of the platform i , which is the observation of ξ_i will cost the decision maker a price of f_i . Then the decision $y(\xi) \in \mathbb{R}^n$ is selected subject to the field size, flow conservation, platform capacity and network structure, which could be represented as $Ax + By(\xi) \leq b(\xi)$, where $B \in \mathbb{R}^{m \times n}$ and at a cost of $c^T y(\xi)$. The decision maker is to find the function $y \in F_{k,n}$ so as to minimize the cost or maximize the profit. Therefore, the decision problem can be formulated in the following general form:

$$\begin{aligned}
& \min f^T x + E(c^T y(\xi)) \\
& s.t. \quad x \in \mathbb{Z}^k, y \in F_{k,n} \\
& \quad \left. \begin{aligned} Ax + d + By(\xi) &\leq h(\xi) \\ y(\xi) &= y(x \bullet \xi) \end{aligned} \right\} \forall \xi \in \Xi,
\end{aligned}$$

where $y(\xi) \in \mathbb{R}^n$ denotes the decision/strategy with respect to the stochastic variable. x is a binary decision vector for construction of pipelines and platforms, with x_i forcing the unobserved variable ξ_i equals to 0 and hence has no effect on the strategy function y . Let Ξ denotes a compact polyhedral subset of $\{\xi \in \mathbb{R}^k : \xi_1 = 1\}$, which will enforce that the affine functions of the non-degenerate uncertain parameters could be represented in a compact way as linear functions of $\xi = (\xi_1, \dots, \xi_k)$.

Approximation

To approximate the one-stage stochastic problem solution, we use a linear assumption of the underlying data of the form

$$\begin{aligned}
y(\xi) &= Y\xi \quad Y \in \mathbb{R}^{n \times k}, \\
h(\xi) &= H\xi \quad H \in \mathbb{R}^{m \times k}.
\end{aligned}$$

This assumption will reduce the admissible decisions/strategies to those that are presented as affine dependence. Then the original stochastic problem is converted to a semi-infinite type as it includes only a finite number of variables but an infinite number of constraints parameterized by $\xi \in \Xi$:

$$\min f^T x + E(c^T Y\xi),$$

$$\begin{aligned}
s.t. \quad & x \in \mathbb{Z}^k, Y \in \mathbb{R}^{n \times k} \\
& \left. \begin{aligned} & Ax + d + BY\xi \leq H\xi \\ & y(\xi) = y(x \bullet \xi) \end{aligned} \right\} \forall \xi \in \Xi.
\end{aligned}$$

Note that the last constraint in the problem can be restated as

$$|Y_{ij}| \leq Mx_j \quad i = 1, \dots, n, \quad j = 1, \dots, k.$$

This set of constraints hold that if ξ_j is not observed, then the decision/strategy

$y(\xi) \in \mathbb{R}^n$ should be independent of ξ_j . But M should be large enough to make sure that

Y_{ij} is unaffected when $x_j = 1$.

The support of the probability measure P could be represented as the form

$$\Xi = \{\xi \in \mathbb{R}^k : W\xi \geq v\}.$$

Proposition: For any $\phi \in \mathbb{R}^k$, the following statements are equivalent:

- (i) $\phi^T \xi \geq Ax + d$ for all $\xi \in \Xi$, where $\Xi = \{\xi \in \mathbb{R}^k : W\xi \geq v\}$;
- (ii) $\exists \lambda \in \mathbb{R}^l$ with $\lambda \geq 0$, $W^T \lambda = \phi$, and $v^T \lambda \geq Ax + d$.

Proof: Using the duality properties of mixed integer linear programming, we have

$$\phi^T \xi \geq Ax + d \text{ for all } \xi \in \Xi, \text{ where } \Xi = \{\xi \in \mathbb{R}^k : W\xi \geq v\}$$

$$\Leftrightarrow \min_{\xi \in \mathbb{R}^k} \{\phi^T \xi : W\xi \geq v\} \geq Ax + d$$

$$\Leftrightarrow \max_{\lambda \in \mathbb{R}^l} \{v^T \lambda : W^T \lambda = \phi, \lambda \geq 0\} \geq Ax + d$$

$$\Leftrightarrow \exists \lambda \in \mathbb{R}^l \text{ with } \lambda \geq 0, \quad W^T \lambda = \phi, \text{ and } v^T \lambda \geq Ax + d.$$

Then the original problem can be reformulated as

$$\min f^T x + c^T Y E(\xi),$$

$$\begin{aligned}
s.t. \quad & x \in \mathbb{Z}^k, Y \in \mathbb{R}^{n \times k}, \Lambda \in \mathbb{R}^{m \times l}, \\
& \Lambda W + BY = H, \\
& \Lambda v - Ax - d \geq 0, \\
& \Lambda \geq 0.
\end{aligned}$$

The above approximation formulation of one-stage stochastic programming can be solved efficiently as a mixed integer binary program. Its size grows polynomially with k, m, n and l , which are the size of the original problem and the number of constraints in the underlying uncertainty set Ξ . The resulting solution is a conservative approximation of the original problem.

Numerical Example

Now considering one-stage decision making in region A of figure 3-1, the decision maker is to maximize the profit of gas production. He is to make decisions only once based on the stochastic information. For simplification, constraints on production rate may be omitted. Instead, the size of the gas field will place a limit on the amount of gas extracted. The model is generalized as follows:

$$\max E \left[c_g \sum_{l \in L^+(o)} y_f^l(\xi) - \sum_{p \in P} c_i^p x^p - c_c^p y_c^p(\xi) - c_e^p y_e^p(\xi) - \sum_{l \in L} c_i^l x^l \right],$$

$$\begin{aligned}
s.t. \quad & y_e^p(\xi) \leq \xi^p \quad \forall p \in P, \\
& 0 \leq y_e^p(\xi) \leq r_p \quad \forall p \in P, \\
& y_e^p(\xi) + \sum_{l \in L^+(p)} y_f^l(\xi) \geq \sum_{l \in L^-(p)} y_f^l(\xi) \quad \forall p \in P, \\
& \sum_{l \in L^-(p)} y_f^l(\xi) \leq y_c^p(\xi) \quad \forall p \in P, \\
& 0 \leq y_f^l(\xi) \leq Mx^l \quad \forall l \in L, \\
& 0 \leq y_c^p(\xi) \leq Mx^p \quad \forall p \in P, \\
& x^l, x^p \in \{0,1\} \quad \forall l \in L, p \in P, \\
& y_f^l \in F_{p,l}, y_e^p \in F_{p,p}, y_c^p \in F_{p,p}.
\end{aligned}$$

As for region A, three platforms and five pipelines are considered to be built at the very beginning. Once a platform is built, the corresponding field size is revealed. The binary variable will enforce the decision rules not relying on unexploited gas fields.

Suppose the input parameters of the problem are summarized in table 4-1.

Table 4-1. Input parameters for one-stage.

ξ^l	Gas field size. Random variable: uniform distributed $U(0,20), U(0,10), U(0,10)$
c_g	Unit price for gas 2
c_c^l	Unit capacity expansion cost for production platform p (0.2, 0.2, 0.2)
c_e^l	Unit gas extraction cost for platform p (0.1, 0.1, 0.1)

c_i^l	Cost for building pipeline l (2, 1, 3, 1, 5)
c_i^p	Cost for building platform p (4, 2, 2)

A detailed formulation of region A can therefore be formulated as

$$\begin{aligned}
 \max E & \left[c_g (y_f^{l_2}(\xi) + y_f^{l_4}(\xi)) - \sum_{p=p_1}^{p_3} c_i^p x^p - c_c^p y_c^p(\xi) - c_e^p y_e^p(\xi) - \sum_{l=l_1}^{l_5} c_i^l x^l \right], \\
 \text{s.t. } & y_e^p(\xi) \leq \xi^p \quad \forall p \in \{p_1, p_2, p_3\}, \\
 & y_e^{p_1}(\xi) \geq y_f^{l_1}(\xi) + y_f^{l_3}(\xi), \\
 & y_e^{p_2}(\xi) + y_f^{l_1}(\xi) \geq y_f^{l_2}(\xi) + y_f^{l_5}(\xi), \\
 & y_e^{p_3}(\xi) + y_f^{l_3}(\xi) + y_f^{l_5}(\xi) \geq y_f^{l_4}(\xi), \\
 & y_c^{p_1}(\xi) \geq y_f^{l_1}(\xi) + y_f^{l_3}(\xi), \\
 & y_c^{p_2}(\xi) \geq y_f^{l_2}(\xi) + y_f^{l_5}(\xi), \\
 & y_c^{p_3}(\xi) \geq y_f^{l_4}(\xi), \\
 & 0 \leq y_f^l(\xi) \leq Mx^l \quad \forall l \in \{l_1, l_2, l_3, l_4, l_5\}, \\
 & 0 \leq y_c^p(\xi) \leq Mx^p \quad \forall p \in \{p_1, p_2, p_3\}, \\
 & x^l, x^p \in \{0, 1\} \quad \forall l \in \{l_1, l_2, l_3, l_4, l_5\}, \forall p \in \{p_1, p_2, p_3\}, \\
 & y_f^l \in F_{3,5}, y_e^p \in F_{3,3}, y_c^p \in F_{3,3}.
 \end{aligned}$$

Following the approximation steps discussed, under the assumption of linear decision strategy and underlying data, we have

$$y(\xi) = Y\xi \quad Y \in \mathbb{R}^{11 \times 3},$$

$$h(\xi) = H\xi \quad H \in \mathbb{R}^{9 \times 3}.$$

The underlying uncertainty set of stochastic variable ξ can also be represented as

$$\Xi = \{\xi \in \mathbb{R}^3 : W\xi \geq v\}.$$

Together with the constraints that enforce the decision/strategy $y(\xi)$ should be independent of any unobserved ξ_j . Then the original problem is reformulated as the following standard mixed integer problem:

$$\max E \left[c_g (Y_2^f E(\xi) + Y_4^f(\xi)) - \sum_{p=p_1}^{p_3} c_i^p x^p - \sum_{l=l_1}^{l_5} c_i^l x^l - c_c^p Y^c E(\xi) - c_e^p Y^e E(\xi) \right],$$

$$s.t. \quad x^p \in \mathbb{Z}^3, x^l \in \mathbb{Z}^5, Y \in \mathbb{R}^{11 \times 3}, \Lambda \in \mathbb{R}^{9 \times 6},$$

$$\Lambda W + BY = H,$$

$$\Lambda v \geq 0,$$

$$\Lambda \geq 0,$$

$$0 \leq Y_i^f \leq Mx^{l_i} \quad i = 1, 2, \dots, 5,$$

$$0 \leq Y_i^e, Y_i^c \leq Mx^{p_i} \quad i = 1, 2, 3,$$

$$0 \leq Y_i^c \leq Mx^{p_i} \quad i = 1, 2, 3,$$

$$0 \leq Y_{ij}^f, Y_{ij}^e, Y_{ij}^c \leq Mx^{p_j} \quad \forall i, j = 1, 2, 3.$$

where

$$B = \begin{pmatrix} B_f & B_e & B_c \end{pmatrix}, \quad Y = \begin{bmatrix} Y_f \\ Y_e \\ Y_c \end{bmatrix},$$

$$B_f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & -1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad B_e = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_c = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad v = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -20 \\ -10 \\ -10 \end{bmatrix}.$$

Solving the above problem using GAMS, we obtain the following result,

$$Y_f = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad Y_c = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Y_e = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$x^p = (1, 1, 1),$$

$$x^l = (1, 1, 0, 1, 0),$$

$$\text{objective function } z = 20.$$

According to Y_f , we can infer that flow goes through pipeline 1 equals to the size of gas field 1, flow of pipeline 2 equals to the size of gas fields 1 and 2 and pipeline 4

delivers the gas from gas field 3. Besides, pipeline 3 and 5 are not constructed in this problem. From Y_c , we can tell that the capacity of each gas field is enough to deliver the gas to the following pipeline. Also, Y_e indicates that all of the gas within each gas well is exploited to obtain the maximize profit.

The results are intuitive. The decision maker is to maximize the profit of the project in one calendar year. Thus, he will need to extract all the gas from each gas field where platforms are built. To transport gas from each platform to the central pipeline, pipelines will be built and platforms will be expanded to a capacity that will pump all the incoming gas to outgoing pipelines. In other words, no gas will be wasted because of insufficient of transportation.

From the result, we can infer that pipelines 3 and 5 are not constructed. This is because pipelines 3 and 5 are relatively expensive as compared to other pipelines, and capacity expansion cost is not high enough for the company to build extra pipelines to transport the gas another way. Also, notice that all the gas fields are exploited as $x^p = (1, 1, 1)$.

Under this scenario, the result is the same with a static problem with the size of each gas field fixed and equal to the expectation of each, which also agrees with Monte Carlo sampling. The problem is bounded and solved efficiently, which obviously is an advantage over Monte Carlo sampling.

Chapter 5

Multi-stage Infrastructure and Production Planning

Model

The dynamic infrastructure and production planning problem in this section is considered in a way that a decision maker makes sequential investment and operational decisions, and obtain observation of the uncertainty parameters $\xi = (\xi_1, \dots, \xi_k)$, which are still defined on the probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathcal{P})$, over a finite planning horizon $T := \{1, \dots, t\}$. Such problem can be formalized as

$$\begin{aligned}
 & \min E\left(\sum_{t \in T} f_t^T x_t(\xi) + c_t^T y_t(\xi)\right), \\
 & s.t. \quad x_t \in F_{k,k}, y_t \in F_{k,n} \quad \forall t \in T, \\
 & \left. \begin{aligned}
 & \sum_{\tau=1}^t A_{t\tau} x_\tau(\xi) + B_{t\tau} y_\tau(\xi) \leq h_t(\xi) \\
 & x_t(\xi) \in Z_t \\
 & x_t(\xi) \geq x_{t-1}(\xi) \\
 & x_t(\xi) = x_t(x_{t-1}(\xi) \bullet \xi) \\
 & y_t(\xi) = y_t(x_{t-1}(\xi) \bullet \xi)
 \end{aligned} \right\} \forall \xi \in \Xi, t \in T,
 \end{aligned}$$

Where $y_t(\xi)$ is a vector that denotes the decision rules/strategies at time t , and $x_t(\xi)$ is an adaptive decision variable that encodes binary information of construction up to time t , which is dependent on the uncertain gas field reserve ξ . The set Z_t , which the

adaptive decision variable belongs to is a subset of $\{0,1\}^k$, as it may include constraints which enforce the order of the gas field reserve revealed. For example, a platform can only be built at a certain gas field after another platform is constructed or certain pipelines can be built only after certain stages. If ξ_i is observed and included in the information base at time t , then $x_{t,i}(\xi) = 1$, which will also incur a cost of $f_{t,i}$ and another term $B_{t\tau}y_\tau(\xi)$ in the time t constraint. The constraint $x_t(\xi) \geq x_{t-1}(\xi)$ will enforce that the construction will not be removed, and thus $x_{t,i}(\xi)$ is monotonous, which will stay on 1. The last two constraints in the formulation enforce non-anticipativity, which restrict the decision strategies $y_t(\xi)$ to only depend on gas fields information obtained up to time $t-1$.

The above type of problem involves a multi-stage dynamic programming with adaptive decision rules/strategies and binary recourse variables and is shown to be computationally intractable. To approximate a conservative solution, linear assumptions and partition of uncertainty set are therefore necessary.

Approximation

Compared to the one-stage production planning model, the multi-stage is more complicated and expensive in computing. Past research on multi-stage stochastic programming has studied the approximation of stochastic programming with continuous recourse variables, of which conservative solutions could be obtained by linear decision rules (Ben-Tal et al. 2004). Also, finite adaptability, which is the middle ground of

complete adaptability where the decision-maker has arbitrary adaptability to the exact realization of the uncertainty and static robust formulation where the decision-maker has no information on the realization of the uncertainty, has also proved tractable and efficient when solving multi-stage stochastic programming (Bertsimas, D., and Caramanis, C, 2010). The idea of the approach is partitioning the uncertainty space and receiving information about the realization of the uncertainty, which provides an opportunity to trade off computing expense with optimality. Based on this idea, Vayanos et al. (2011) solve the stochastic programming problem with endogenous uncertainty by approximating the binary decision rules that are piecewise constant and real-valued decisions that are piecewise linear with respect to a pre-selected partition set.

Let Ξ_s denotes the subset of the partition of the uncertainty set

$$\Xi_s := \left\{ \xi \in \Xi : a_{s_i-1}^i \leq \xi_i < a_{s_i}^i, i = 1, \dots, k \right\},$$

$$\text{where } s \in S := \times_{i=1}^k \{1, \dots, r_i\} \subseteq \mathbb{N}^k,$$

which separate the original uncertainty set Ξ into $(r_i)^k$ subset by breaking along the ξ_i axis into r_i parts.

Thus, the piecewise constant binary decision rule has the form

$$x_t(\xi) = \sum_{s \in S} I_{\Xi_s}(\xi) x_t^s,$$

where $x_t^s \in \{0, 1\}^k$, $s \in S$, $t \in T$ and I_{Ξ_s} denotes the indicator function of Ξ_s .

Similarly, real-valued decisions can be approximated by piecewise linear decision rules of the form

$$y_t(\xi) = \sum_{s \in S} \mathbf{I}_{\Xi_s}(\xi) Y_t^s \xi,$$

where $Y_t^s \in \mathbb{R}^{n_t \times k}$, $s \in S$, $t \in T$.

Under the above assumptions, the non-anticipativity constraints

$$x_t(\xi) \geq x_t(x_{t-1}(\xi) \bullet \xi),$$

$$y_t(\xi) \geq y_t(x_{t-1}(\xi) \bullet \xi),$$

can be re-expressed as

$$\forall \xi, \xi', t : x_{t-1}(\xi) \bullet \xi \geq x_{t-1}(\xi') \bullet \xi',$$

$$x_t(\xi) = x_t(\xi'),$$

$$y_t(\xi) = y_t(\xi').$$

Substituting the assumptions into the above equations, we have:

$$\forall s, s', t : x_{t-1}^s \bullet s \geq x_{t-1}^{s'} \bullet s',$$

$$x_t^s = x_t^{s'},$$

$$Y_t^s = Y_t^{s'},$$

and

$$\forall i, j, s, t \quad |Y_{t,ij}^s| \leq M x_{t-1,j}^s.$$

Note that non-anticipativity across distinct subsets of the partition is enforced in the former part of the constraints while a restriction within each subset is placed in the latter part.

Vayanos et al. (2009) then reformulate the above constraints to reduce the notational overhead by suppressing the domain of the variables as follows:

$$\forall j, j', s, s', t : s_{-j} = s'_{-j},$$

$$\left| x_{t,j'}^s - x_{t,j'}^{s'} \right| \leq x_{t-1,j}^s,$$

$$\left| Y_{t,ij'}^s - Y_{t,ij'}^{s'} \right| \leq M x_{t-1,j}^s \quad \forall i,$$

and

$$\forall i, j, s, t \quad \left| Y_{t,ij}^s \right| \leq M x_{t-1,j}^s.$$

Therefore, the original problem is reformulated as

$$\min E(\sum_{t \in T} f_t^T x_t(\xi) + c_t^T y_t(\xi)),$$

$$s.t. \quad x_t \in F_{k,k}, y_t \in F_{k,n} \quad \forall t \in T,$$

$$\left. \begin{aligned} \sum_{\tau=1}^t A_{t\tau} x_{\tau}(\xi) + B_{t\tau} y_{\tau}(\xi) &\leq h_t(\xi) \\ x_t(\xi) &\in Z_t \\ x_t(\xi) &\geq x_{t-1}(\xi) \end{aligned} \right\} \forall \xi \in \Xi, t \in T,$$

$$\left. \begin{aligned} \left| x_{t,j'}^s - x_{t,j'}^{s'} \right| &\leq x_{t-1,j}^s \\ \left| Y_{t,ij'}^s - Y_{t,ij'}^{s'} \right| &\leq M x_{t-1,j}^s \quad \forall i \end{aligned} \right\} \forall j, j', s, s', t : s_{-j} = s'_{-j},$$

$$\left| Y_{t,ij}^s \right| \leq M x_{t-1,j}^s \quad \forall i, j, s, t.$$

where

$$x_t(\xi) = \sum_{s \in S} I_{\Xi_s}(\xi) x_t^s, \quad y_t(\xi) = \sum_{s \in S} I_{\Xi_s}(\xi) Y_t^s \xi.$$

The partition of the uncertainty set could be represented as the form

$$\Xi_s = \{\xi \in \mathbb{R}^k : W_s \xi \geq v_s\}.$$

We follow similar steps in one-stage stochastic programming. The original problem can be reformulated as a mixed-binary linear programming problem by

substituting the piecewise constant and linear assumptions into every other constraints.

Hence,

$$\begin{aligned}
& \min \sum_{s \in S} p_s \sum_{t \in T} f_t^T x_t^s + c_t^T Y_t^s E(\xi), \\
& s.t. \quad x_t^s \in Z_t, Y_t^s \in \mathbb{R}^{n_t \times k}, \Lambda_t^s \in \mathbb{R}^{m_t \times l_s} \quad \forall s, t, \\
& \left. \begin{aligned}
& \Lambda_t^s W_s + \sum_{\tau=1}^t B_{t\tau} Y_t^s = H_t^s \\
& \Lambda_t^s v_s - \sum_{\tau=1}^t A_{t\tau} x_t^s \geq 0 \\
& \Lambda_t^s \geq 0
\end{aligned} \right\} \forall s, t, \\
& x_t^s \geq x_{t-1}^s \quad \forall s, t, \\
& \left. \begin{aligned}
& |x_{t,j}^s - x_{t,j'}^{s'}| \leq x_{t-1,j}^s \\
& |Y_{t,ij}^s - Y_{t,ij'}^{s'}| \leq M x_{t-1,j}^s \quad \forall i
\end{aligned} \right\} \forall j, j', s, s', t : s_{-j} = s'_{-j}, \\
& |Y_{t,ij}^s| \leq M x_{t-1,j}^s \quad \forall i, j, s, t.
\end{aligned}$$

The new formulation of multi-stage stochastic programming has the form of a standard mixed-binary linear programming. Its size is bounded by the size of the original problem, the partition of the uncertainty set and the number of constraints in each underlying uncertainty set Ξ_s . The resulting solution is a conservative approximation of the original problem.

Numerical Example

Now, we consider the problem in a more comprehensive way. Assume the company plans taking on the project for a ten-year period, during which time the decision

maker needs to organize the gas extraction and transportation process. The goal is to maximize the net present value of the project.

Different from the one-stage problem, the decision maker now has multiple years to make decisions for gas production. The size of each gas field is not necessarily revealed in the first period, as the decision maker may choose a later time to reduce the cost of building the platform. However, the revenue of the sales of gas will also decrease when the gas is extracted at a later time.

Thus, in the multi-stage shale gas production problem, both strategy and binary variables which represent the construction of platform and pipeline will be dependent of the revelation of the size of gas wells. As discussed previously, the multi-year gas production model can be formulated as follows:

$$\max \sum_{t \in \Gamma} d_t E \left[c_g \sum_{l \in L^+(o)} y_{f,t}^l(\xi) - \sum_{l \in L} c_i^l (x_t^l(\xi) - x_{t-1}^l(\xi)) \right. \\ \left. - \sum_{p \in P} c_i^p (x_t^p(\xi) - x_{t-1}^p(\xi)) - c_c^p y_{c,t}^p(\xi) - c_e^p y_{e,t}^p(\xi) \right],$$

$$\begin{aligned}
s.t. \quad & \sum_{t \in \Gamma} y_{e,t}^p(\xi) \leq \xi^p \quad \forall p \in P, t \in \Gamma, \\
& 0 \leq y_{e,t}^p(\xi) \leq r_p \quad \forall p \in P, t \in \Gamma, \\
& y_{e,t}^p(\xi) + \sum_{l \in L^+(p)} y_{f,t}^l(\xi) \geq \sum_{l \in L^-(p)} y_{f,t}^l(\xi) \quad \forall p \in P, t \in \Gamma, \\
& \sum_{l \in L^-(p)} y_{f,t}^l(\xi) \leq \sum_{\tau=1}^t y_{c,\tau}^p(\xi) \quad \forall p \in P, t \in \Gamma, \\
& 0 \leq y_{f,t}^l(\xi) \leq Mx_t^l(\xi) \quad \forall l \in L, t \in \Gamma, \\
& 0 \leq y_{c,t}^p(\xi) \leq Mx_t^p(\xi) \quad \forall p \in P, t \in \Gamma, \\
& x_{t-1}^p(\xi) \leq x_t^p(\xi) \quad \forall p \in P, t \in \Gamma, \\
& x_{t-1}^l(\xi) \leq x_t^l(\xi) \quad \forall l \in L, t \in \Gamma, \\
& x_t^l, x_t^p \in \{0,1\} \quad \forall l \in L, p \in P, \\
& y_f^l \in F_{p,l}, y_e^p \in F_{p,p}, y_c^p \in F_{p,p}.
\end{aligned}$$

To make a comparison between multi-stage stochastic programming on the shale gas production project, we assume similar input parameters as for the one-stage problem. Thus, input parameters are shown as in table 5-1. Notice that an upper limit is placed on the annual production rate, forcing the decision maker taking account of longer period.

Table 5-1. Input parameters for multi-stage.

ξ^l	Gas field size. Random variable: uniform distributed $U(0,20), U(0,10), U(0,10)$
r_p	Maximum annual production rate (2, 2, 2)
d_t	Discount factor in year t

	$\frac{1}{1.01^{t-1}}$
c_g	Unit price for gas
	2
c_c^P	Unit capacity expansion cost for production platform P
	(0.2, 0.2, 0.2)
c_e^P	Unit gas extraction cost for platform P
	(0.1, 0.1, 0.1)
c_i^l	Cost for building pipeline l
	(2, 1, 3, 1, 5)
c_i^P	Cost for building platform P
	(4, 2, 2)

Now considering the multi-stage decision making problem within region A, the problem can be described as follows:

$$\max \sum_{t=1}^{10} d_t E \left[\begin{array}{l} c_g (y_{f,t}^{l_2}(\xi) + y_{f,t}^{l_4}(\xi)) - \sum_{l=l_1}^{l_5} c_i^l (x_t^l(\xi) - x_{t-1}^l(\xi)) \\ - \sum_{p=p_1}^{p_3} c_i^p (x_t^p(\xi) - x_{t-1}^p(\xi)) - c_c^p y_{c,t}^p(\xi) - c_e^p y_{e,t}^p(\xi) \end{array} \right],$$

$$s.t. \quad \forall t = 1, 2, \dots, 10,$$

$$\sum_{\tau=1}^t y_{e,\tau}^p(\xi) \leq \xi^p \quad \forall p \in \{p_1, p_2, p_3\},$$

$$0 \leq y_{e,t}^p(\xi) \leq r_p \quad \forall p \in \{p_1, p_2, p_3\},$$

$$y_{e,t}^{p_1}(\xi) \geq y_{f,t}^{l_1}(\xi) + y_{f,t}^{l_3}(\xi),$$

$$y_{e,t}^{p_2}(\xi) + y_{f,t}^{l_1}(\xi) \geq y_{f,t}^{l_2}(\xi) + y_{f,t}^{l_5}(\xi),$$

$$y_{e,t}^{p_3}(\xi) + y_{f,t}^{l_3}(\xi) + y_{f,t}^{l_5}(\xi) \geq y_{f,t}^{l_4}(\xi),$$

$$\sum_{\tau=1}^t y_{c,\tau}^{p_1}(\xi) \geq y_{f,t}^{l_1}(\xi) + y_{f,t}^{l_3}(\xi),$$

$$\sum_{\tau=1}^t y_{c,\tau}^{p_2}(\xi) \geq y_{f,t}^{l_2}(\xi) + y_{f,t}^{l_5}(\xi),$$

$$\sum_{\tau=1}^t y_{c,\tau}^{p_3}(\xi) \geq y_{f,t}^{l_4}(\xi),$$

$$0 \leq y_{f,t}^l(\xi) \leq Mx_t^l(\xi) \quad \forall l \in \{l_1, l_2, l_3, l_4, l_5\},$$

$$0 \leq y_{c,t}^p(\xi) \leq Mx_t^p(\xi) \quad \forall p \in \{p_1, p_2, p_3\},$$

$$x_{t-1}^p(\xi) \leq x_t^p(\xi) \quad \forall p \in \{p_1, p_2, p_3\},$$

$$x_{t-1}^l(\xi) \leq x_t^l(\xi) \quad \forall l \in \{l_1, l_2, l_3, l_4, l_5\},$$

$$x_t^l, x_t^p \in \{0, 1\} \quad \forall l \in \{l_1, l_2, l_3, l_4, l_5\}, \forall p \in \{p_1, p_2, p_3\},$$

$$y_{f,t}^l \in F_{3,5}, y_{e,t}^p \in F_{3,3}, y_{c,t}^p \in F_{3,3}.$$

To approximate the decision rule, we partition the uncertainty set ξ^p into some preselected subsets. Specifically, each $\xi^{p_i} \quad \forall i = 1, 2, 3$ is divided into two parts. With each part denoted by Ξ_s , we have a total of eight subsets with the same probability. The assumption that gas fields follow an uniform distribution makes the partition

straightforward and easy to implement. Table 5-2 shows the partition of the entire uncertainty set.

Table 5-2. Partition of uncertainty set.

Subset Ξ_s ($P_s = \frac{1}{8}$)	Gas fields $\xi^{p_1}, \xi^{p_2}, \xi^{p_3}$
$s = 1$	$U(0,10), U(0,5), U(0,5)$
$s = 2$	$U(10,20), U(0,5), U(0,5)$
$s = 3$	$U(0,10), U(5,10), U(0,5)$
$s = 4$	$U(10,20), U(5,10), U(0,5)$
$s = 5$	$U(0,10), U(0,5), U(5,10)$
$s = 6$	$U(10,20), U(0,5), U(5,10)$
$s = 7$	$U(0,10), U(5,10), U(5,10)$
$s = 8$	$U(10,20), U(5,10), U(5,10)$

Based on the above partition, we approximate the measurement decisions by binary-valued decision rules that are piecewise constant and the real-valued decisions by decision rules that are piecewise constant.

The binary-valued decisions are approximated in the form.

$$x_t^l(\xi) = \sum_{s \in S} I_{\Xi_s}(\xi) x_t^{l,s},$$

$$x_t^p(\xi) = \sum_{s \in S} I_{\Xi_s}(\xi) x_t^{p,s},$$

where

$$x_t^{l,s} \in \{0,1\}^5, \quad s \in S, \quad t \in \Gamma,$$

$$x_t^{p,s} \in \{0,1\}^3, \quad s \in S, \quad t \in \Gamma,$$

and I_{Ξ_s} denotes the indicator function of Ξ_s .

Similarly, real valued decisions can be approximated as the form

$$y_{f,t}^l(\xi) = \sum_{s \in S} I_{\Xi_s}(\xi) Y_{f,t}^{l,s} \xi,$$

$$y_{e,t}^p(\xi) = \sum_{s \in S} I_{\Xi_s}(\xi) Y_{e,t}^{p,s} \xi,$$

$$y_{c,t}^p(\xi) = \sum_{s \in S} I_{\Xi_s}(\xi) Y_{c,t}^{p,s} \xi,$$

$$\text{where } Y_{f,t}^{l,s} \in \mathbb{R}^{5 \times 3}, Y_{e,t}^{p,s} \in \mathbb{R}^{3 \times 3}, Y_{c,t}^{p,s} \in \mathbb{R}^{3 \times 3}.$$

Apart from the constraints we already had in the original formulation, the reformulation requires additional non-anticipativity constraints, which further ensure that decision variables are independent of unrevealed gas fields. Under the partition above, such constraints can be represented as

$$x_t^s(\xi) = x_t^s(x_{t-1}^s(\xi) \bullet \xi),$$

$$y_t^s(\xi) = y_t^s(x_{t-1}^s(\xi) \bullet \xi),$$

which are equivalent to

$$\xi, \xi' \in \Xi, \quad s, s' \in S, \quad t \in \{1, 2, \dots, 10\}, \quad j, j' \in \{1, 2, 3\},$$

$$\left. \begin{aligned} & \left| x_{t,j'}^{p,s} - x_{t,j'}^{p,s'} \right| \leq x_{t-1,j}^{p,s} \\ & \left| Y_{e,t,ij'}^{p,s} - Y_{e,t,ij'}^{p,s'} \right| \leq M x_{t-1,j}^{p,s} \quad \forall i = 1, 2, 3 \\ & \left| Y_{c,t,ij'}^{p,s} - Y_{c,t,ij'}^{p,s'} \right| \leq M x_{t-1,j}^{p,s} \quad \forall i = 1, 2, 3 \\ & \left| Y_{f,t,ij'}^{l,s} - Y_{f,t,ij'}^{l,s'} \right| \leq M x_{t-1,j}^{p,s} \quad \forall i = 1, 2, 3, 4, 5 \end{aligned} \right\} \forall l, j, j', s, s', t : s_{-j} = s'_{-j},$$

$$\left. \begin{aligned} |Y_{f,t,ij}^{l,s}| &\leq Mx_{t-1,j}^{p,s} \\ |Y_{e,t,ij}^{p,s}| &\leq Mx_{t-1,j}^{p,s} \\ |Y_{c,t,ij}^{p,s}| &\leq Mx_{t-1,j}^{p,s} \end{aligned} \right\} \forall i, j, s, t,$$

where $s_{-j} := (s_1, \dots, s_{j-1}, s_{j+1}, \dots, s_k) \in \mathbb{R}^{k-1}$.

After substituting the linear assumption and taking into account the non-anticipativity constraints, the optimization problem becomes

$$\min \sum_{s=s_1}^{s_8} p_s \sum_{t=1}^{10} d_t \left[\begin{aligned} &c_g (Y_{f,t}^{l_2,s} E(\xi_{\Xi_s}) + Y_{f,t}^{l_4,s} E(\xi_{\Xi_s})) - \sum_{l=l_1}^{l_5} c_i^l (x_t^{l,s} - x_{t-1}^{l,s}) - \\ &\sum_{p=p_1}^{p_3} c_i^p (x_t^{p,s} - x_{t-1}^{p,s}) - c_c^p Y_{c,t}^{p,s} E(\xi_{\Xi_s}) - c_e^p Y_{e,t}^{p,s} E(\xi_{\Xi_s}) \end{aligned} \right],$$

$$s.t. \quad x_t^s \in Z_t, Y_{f,t}^{l,s} \in \mathbb{R}^{5 \times 3}, Y_{e,t}^{p,s} \in \mathbb{R}^{3 \times 3}, Y_{c,t}^{p,s} \in \mathbb{R}^{3 \times 3}, \Lambda_t^s \in \mathbb{R}^{12 \times 6},$$

$$\left. \begin{aligned} &\Lambda_t^s W_s + \sum_{\tau=1}^t B_{t\tau}^f Y_{f,t}^{l,s} + B_{t\tau}^e Y_{e,t}^{p,s} + B_{t\tau}^c Y_{c,t}^{p,s} = H_t^s \\ &\Lambda_t^s v_s \geq d_t^s \\ &\Lambda_t^s \geq 0 \end{aligned} \right\} \forall s, t,$$

$$x_t^{l,s} \geq x_{t-1}^{l,s} \quad \forall s, t,$$

$$x_t^{p,s} \geq x_{t-1}^{p,s} \quad \forall s, t,$$

$$\left. \begin{aligned} &|x_{t,j'}^{p,s} - x_{t,j'}^{p,s'}| \leq x_{t-1,j}^{p,s} \\ &|Y_{e,t,ij'}^{p,s} - Y_{e,t,ij'}^{p,s'}| \leq Mx_{t-1,j}^{p,s} \quad \forall i=1,2,3 \\ &|Y_{c,t,ij'}^{p,s} - Y_{c,t,ij'}^{p,s'}| \leq Mx_{t-1,j}^{p,s} \quad \forall i=1,2,3 \\ &|Y_{f,t,ij'}^{l,s} - Y_{f,t,ij'}^{l,s'}| \leq Mx_{t-1,j}^{p,s} \quad \forall i=1,2,3,4,5 \end{aligned} \right\} \forall l, j, j', s, s', t : s_{-j} = s'_{-j},$$

$$\left. \begin{aligned} |Y_{f,t,ij}^{l,s}| &\leq Mx_{t-1,j}^{p,s} \\ |Y_{e,t,ij}^{p,s}| &\leq Mx_{t-1,j}^{p,s} \\ |Y_{c,t,ij}^{p,s}| &\leq Mx_{t-1,j}^{p,s} \end{aligned} \right\} \forall i, j, s, t,$$

$$\xi, \xi' \in \Xi, \quad s, s' \in S, \quad t \in \{1, 2, \dots, 10\}, \quad j, j' \in \{1, 2, 3\}.$$

The above optimization problem is a standard mixed-integer programming with its size bounded by the size of the original problem, the partition of the uncertainty set and the number of constraints in each underlying uncertainty set Ξ_s . I use GAMS with CPLEX solver to solve the problem. It takes the software 5 seconds to obtain the result. Some optimal results of decision variables are indicated and explained below.

As binary variables in this problem are monotonic, platforms and pipelines will not be deconstructed and only the periods when pipelines and platforms are built are indicated in table 5-3.

Table 5-3. Optimal construction.

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
l_1	NA	1	1	1	NA	1	1	1
l_2	9	1	1	1	8	1	1	1
l_3	3	NA	NA	NA	6	NA	NA	NA
l_4	1	NA	6	9	6	4	1	1
l_5	NA	NA	NA	NA	NA	NA	NA	NA
p_1	1	1	1	1	1	1	1	1

p_2	1	1	1	1	1	1	1	1
p_3	1	3	1	1	2	3	1	1

The numbers in the table indicate the index of period t .

NA means the pipeline/platform is not built under this scenario.

Notice that most of the gas fields are exploited during the first period. This is because the revenue of the gas in the first year is more profitable compared to the cost of building pipelines and platforms. We can also see that different results are obtained under different partition subsets. For instance, pipeline 4 is built in a different time period for most scenarios. Platform 3 is not exploited in the first period within subset s_2 , s_5 and s_6 . Nonetheless, pipeline 5 is never constructed under all scenarios, which agrees with the result of the one-stage gas production model. The difference between each uncertainty subset demonstrates the necessity of partition when approximating the result.

The amounts of gas extracted from each platform is indicated in table 3-6.

Table 5-4. Optimal gas extraction.

	p_1	p_2	p_3
t_1	0.929	0.875	0.453
t_2	0.929	0.875	0.375
t_3	1.018	0.875	0.453
t_4	1.018	0.875	0.641
t_5	1.018	0.875	0.641
t_6	1.143	0	0.437

t_7	1.143	0	0.438
t_8	0.964	0.104	0.437
t_9	0.964	0.229	0.375
t_{10}	0.875	0.229	0.375
Total	10.001	4.937	4.625

From the table, we can tell that the total amount of gas extracted from each platform is close to the expectation of each field size. Other factors that might affect the amount of gas extracted include the capacity of the platforms that the gas transports, the construction cost of the pipeline and the layout of the network. Notice that the limitation of gas production rate is not a deciding factor in this case.

The amount of capacity expanded during each period is indicated in table 5-5.

Table 5-5. Optimal capacity expansion.

	p_1	p_2	p_3
t_1	0.929	1.804	0.453
t_2	0	0	0
t_3	0.089	0	0.089
t_4	0	0	0.187
t_5	0	0	0
t_6	0.125	0	0.375

t_7	0	0	0
t_8	0	0.104	0
t_9	0	0.125	0.125
t_{10}	0	0	0
Total	1.143	2.033	1.229

It is obvious that decision maker would rather choose a later period to increase capacity as the cost goes down period by period due to the discount factor. Platform 2 has a relatively high capacity, because the gas from platform 1 needs to go through platform 2 to reach the central pipeline. Another reason is that in most cases pipeline 3 is not built due to the high cost which makes pipeline 1 the only way to transport gas from platform 1.

The amounts of capacity expanded during each period are indicated in table 5-6.

Table 5-6. Optimal gas flow.

	l_1	l_2	l_3	l_4	l_5
t_1	0.929	1.804	0	0.453	0
t_2	0.929	1.804	0	0.375	0
t_3	0.929	1.804	0.089	0.542	0
t_4	0.929	1.804	0.089	0.73	0
t_5	0.929	1.804	0.089	0.73	0

t_6	0.929	0.929	0.214	0.652	0
t_7	0.929	0.929	0.214	0.652	0
t_8	0.75	0.854	0.214	0.652	0
t_9	0.75	0.979	0.214	0.589	0
t_{10}	0.75	0.979	0.125	0.5	0

Since pipeline 5 is never built under any scenarios, the flow through pipeline 5 will always be zero. However, pipeline 3 is built in t_3 under subset s_1 and in t_6 under subset s_5 . Then the flow through pipeline 3 will not be zero as it is calculated as expectation which averages the flow under different scenarios. We also notice that the flow in pipeline 3 goes up in t_6 as pipeline 3 is built under another scenario. This evidence confirms the above explanation.

Finally the objective function is

$$z = 22.817.$$

Compared to the one-stage project, the multi-stage project has a higher net present value which allows the decision maker to earn an extra profit. The conservation approximation obtained lies in-between the situation when the decision-maker has arbitrary adaptability to the exact realization of the uncertainty and the static robust formulation where the decision-maker has no information on the realization of the uncertainty. The approximation could be more accurate if the number of partitions is

increased. However, this will also increase the computing cost as the size of the problem is dependent on the partition.

Chapter 6

Conclusions and Future Work

The shale gas infrastructure and production planning problem is discussed in this paper. The problem is modeled as stochastic programming with endogenous uncertainties. Approximations using the adaptive measurement decisions by piecewise constant functions and the adaptive real-valued decisions by piecewise linear functions of the uncertainties could be applied to obtain conservative solutions.

The decision rule approximation successfully solves the problem with continuously distributed uncertainty parameters. The approximation is considered close to optimal and can be improved by increasing the partition of the uncertainty set. A one-stage numerical is equivalent to a static problem where each field size equals the expectation of the underlying uncertainty parameter. Results of the multi-stage numerical experiment are reasonable. Decision makers are able to obtain more profit by taking on the project for multiple years.

Future work can improve the sophistication of the shale gas infrastructure. A more complicated network with a longer investigating time horizon can be considered to better fit the context of shale gas production. The partition of subset maybe increased if necessary to obtain more accurate solution.

List of References

Barnes, R. J., Linke, P., & Kokossis, A., Optimization of oil-field development production capacity. European Symposium on Computer Aided Process Engineering, 12, 631 (2002).

Ben-Tal, A., Boyd, S., and Nemirovski, A. Extending scope of robust optimization: comprehensive robust counterparts of uncertain problems. Math. Program. 107, 1-2, Ser. B (2006), 63–89.

Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. Adjustable robust solutions of uncertain linear programs. Math. Program. 99, 2, Ser. A (2004), 351–376.

Bertsimas, D., and Caramanis, C. Finite adaptability for linear optimization, IEEE Trans. Automat. Contr., vol. 55, no. 12, pp. 2751–2766, 2010.

Considine, T., and Watson, R. An Emerging Giant: Prospects and Economic Impacts of Developing the Marcellus Shale Natural Gas Play (2009).

Considine, T., and Watson, R., Blumsack, S. The Economic Impacts of the Pennsylvania Marcellus Shale Natural Gas Play: An Update (2010).

Dyer, M., and Stougie, L. Computational complexity of stochastic programming problems. Math. Program. A 106, 3 (2006), 423–432.

Goh, J., and Sim, M., Distributionally robust optimization and its tractable approximations, Oper. Res. vol. 58, no. 4, pp. 902–917, 2010.

Haugen, K. K., A stochastic dynamic programming model for scheduling of offshore petroleum fields with resource uncertainty. *European Journal of Operational Research*, 88, 88 (1996).

Ierapetritou, M. G., Floudas, C. A., Vasantharajan, S., & Cullick, A. S. Optimal location of vertical wells: A decomposition approach. *AIChE Journal*, 45, 844 (1998).

Iyer, R. R., Grossmann, I. E., Vasantharajan, S., & Cullick, A. S. Optimal planning and scheduling of offshore oil field infrastructure investment and operations. *Industrial and Engineering Chemistry Research*, 37, 1380 (1998).

Jonsbraten, T., Optimization models for petroleum field exploitation, Ph.D. dissertation, Norwegian School of Economics and Business Administration (1998).

Kosmidis, V. D., Perkins, J. D., & Pistikopoulos, E. N. A mixed integer optimization strategy for integrated gas/oil production. *European Symposium on Computer Aided Process Engineering*, 12, 697 (2002).

Kuhn, D., Wiesemann, W., and Georghiou, A., Primal and dual linear decision rules in stochastic and robust optimization, *Mathematical Programming*, 2009.

Lin, X., & Floudas, C. A., A novel continuous-time modeling and optimization framework for well platform planning problems. *Optimization and Engineering*, 4, 65–93 (2003).

Meister, B., Clark, J. M. C., & Shah, N., Optimization of oil-field exploitation under uncertainty. *Computers and Chemical Engineering*, 20(ser. B), S1251 (1996).

Ortiz-Gomez, A., Rico-Ramirez, V., & Hernandez-Castro, S., Mixed-integer multi-period model for the planning of oil-field production. *Computers and Chemical Engineering*, 26(4–5), 703 (2002).

Shapiro, A., and Nemirovski, A. On complexity of stochastic programming problems. In *Continuous Optimization: Current Trends and Applications* (2005), V. Jeyakumar and A. Rubinov, Eds., Springer, pp. 111–144.

Vayanos. P., Kuhn, D., and Rustem, B. Decision rules for information discovery in multi-stage stochastic programming. *IEEE CDC-ECC* (2011).