BIO-ENERGY LOGISTICS NETWORK DESIGN UNDER PRICE-BASED SUPPLY AND YIELD UNCERTAINTY

A Dissertation

by

GOKHAN MEMISOGLU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Halit Üster
Co-Chair of Committee,	Justin Yates
Committee Members,	Sıla Çetinkaya
	Neil Geismar
Head of Department,	César O. Malavé

December 2014

Major Subject: Industrial Engineering

ABSTRACT

In this dissertation, we study the design and planning of bio-energy supply chain networks. This dissertation consists of 3 studies that focus on different aspects of bio-energy supply chain systems.

In the first study, we consider planning and design of an extended supply chain for bio-energy networks in an integrated fashion while simultaneously addressing strategic and tactical decisions pertaining to location, production, inventory, and distribution in a multi-period planning horizon setting. For an efficient solution of our model, we suggest a Benders Decomposition based algorithm that can handle realistic size problems for design and analysis purposes. We provide computational results that demonstrate the efficiency of the solution approach on a wide ranging set of problem instances. Furthermore, we develop a realistic case by utilizing data pertaining to the state of Texas and conduct an extensive analysis on the effects of varying input parameters on the design outcomes for a bio-energy supply chain network.

In the second study, we consider a two-stage stochastic problem to model farm-to-biorefinery biomass logistics while designing a policy that encourages farmers to plant biomass energy crops by offering them a unit wholesale price. In the first-stage, the model determines the supply chain network structure as well as the policy parameter, which is the biomass wholesale price offered to farmers. Second-stage problem is to determine the logistical decisions such as transportation, salvaging and out-sourcing. To solve this problem, we propose a solution framework that uses an algorithm based on the L-shaped method along with a Sample Average Approximation (SAA) approach. An extensive case study by varying some of the problem input parameters

is conducted in Texas and the effects on the policy parameter (wholesale price), supply chain network design and expected total system cost are observed.

In the last study, we propose a two-stage stochastic program to model a multi-period biomass-biofuel supply chain system to maximize the expected total system profit. We utilize a similar policy used in the second study to stimulate biomass energy crop production. Our model determines the policy parameter and the supply chain network structure in the first-stage and the tactical decisions for every time period in the second-stage. To solve this problem efficiently, we propose a solution algorithm based on the L-shaped method. Moreover, we also employ SAA approach in our solution methodology to statistically justify our solution quality. A case study is conducted in Texas for different biofuel prices and we analyze changes in the expected system profit the policy parameter and the supply chain network structure. Our case study results indicate that biofuel price needs to be at least \$2.62/gal for the system to have a profit.

DEDICATION

I dedicate this dissertation to my parents...

ACKNOWLEDGMENTS

My Ph.D. journey was a humbling and rewarding process. There are many people I want to recognize and thank for making this journey easier and more enjoyable.

First, I would like to begin by expressing my gratitude to my advisor, Dr. Halit Üster for his continuous support, motivation, understanding and patience throughout my Ph.D. study. Having him as my mentor and with his guidance, I learned a lot and was able to improve myself.

Also, I'd like to thank my Ph.D. committee: Dr. Sıla Çetinkaya, Dr. Neil Geismar and Dr. Justin Yates, for their support, encouragement and insightful comments.

My sincere thanks go to my friends in the Industrial & Systems Engineering Department at Texas A&M University: Abhilasha, Anurag, Arupa, John, Jyotirmoy, Manish, Michelle, Samyukta, Sasha, Su and Sung Ook, for all the difficult, stressful but good times that we spent together that I will always remember and cherish. I also want to thank to my dear friends in College Station: Miray, Engin and their precious daughter Defne; Elif, Gökhan and their beautiful daughter Bade; Eren, Zafer, Cem, Mustafa Gökhan and Dilber, for bringing joy to my life in College Station.

Last but not least, I would like to thank to my family. I know that I was not the easiest child to raise and therefore, first I would like to thank my mother for being very patient and understanding with me. I thank to my father for always showing me his never-ending optimism and goodwill. I'm also extremely grateful to my brother for providing me invaluable advices and for always supporting and protecting me.

There are many other people I could not list here, who stood by me and helped me with the difficulties I encountered during this journey. I'm very grateful to each and every one of them.

This study was supported by USDA through a grant with South Carolina State University, under the project entitled: Optimization and Simulation of Integrated Biomass and Bio-energy Logistics Networks.

NOMENCLATURE

GIS	Geographic Information Systems
LP	Linear Program
MIP	Mixed-integer Program
SP	Stochastic Program
DEP	Deterministic Equivalent Program
SAA	Sample Average Approximation
EPA	U.S. Environmental Protection Agency
DOE	U.S. Department of Energy
EIA	U.S. Energy Information Administration
USDA	U.S. Department of Agriculture
dt	dry ton

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1 Background Information

Worldwide energy consumption has increased dramatically in recent years with the effects of globalization and the economic and industrial advancements in the developing countries like China and India. In 2010, the world energy consumption increased by 5.6%, which is the largest since 1973. Recently, China has surpassed the U.S. and now is leading the world in energy consumption using more than 20% of the global energy. As the technology and the needs of the societies increase, the upward trend in energy consumption will likely to continue. In fact, United States Energy Information Administration [62] projects worldwide energy consumption to increase by 53% between 2008 and 2035. Moreover, it is also estimated that the transportation sector's energy use, which constitutes an important portion of the total energy consumption, increases 1.4% each year from 2008 to 2035. In the U.S., transportation sector is the second largest energy consumption sector after electric power production. United States Energy Information Administration [64] estimates that 28% of the total energy consumption in the U.S. goes to the transportation sector and 93% of this energy consumption is satisfied by petroleum products. Unfortunately, only 4% of the energy consumed in transportation sector is satisfied using renewable energy sources.

The world's leading fuel was oil in 2010, contributing more than 33.6% of the global energy consumption. The global oil consumption grew by 3.1% or equivalently, 2.7 million barrels a day, but the oil production could not keep up with this consumption level and grew only by 2.2% or equivalently, 1.8 million barrels a day [49]. As it is stated in the report published by United States Energy Informa-

tion Administration [64], oil, natural gas and coal are the main energy sources, which constitute around 82% of the total energy production in the U.S. These are fossil fuel sources that produce greenhouse and/or other harmful gases and they are not renewable. Therefore, they are not sustainable in the long-run to secure the energy risks that the world has started facing and they damage the environment by causing pollution. These create environmental risks and energy sustainability issues for our future. Hence, new and alternative ways to generate energy have become one of the top priorities.

Although fossil fuels remain the dominant source of energy, renewable energy is projected to be the fastest growing source of primary energy over the next 25 years [62]. Among these renewable energy sources bio-energy, which is an energy produced from organic matter (biomass) is seen as a substantial future energy source. One of the major advantages of bio-energy over other renewable energy sources such as wind and solar energy, is that biomass can be stored and used on demand. In fact, people have been using bio-energy ever since they started burning wood to provide heat to stay warm and to cook. Nowadays, bio-energy provides more than just heat but electricity, fuel and other types of energies that we need in our daily lives. Although the conversion of biomass for heat and power generation is the most common form of bio-energy, biofuels i.e., liquid fuels produced from biomass, are becoming increasingly popular especially in the transportation sector. The world's biofuel production grew by 13.8% in 2010, mostly by the contributions of the U.S. and Brazil that are the largest corn and sugarcane producers in the world, respectively [49]. Ethanol and biodiesel are the two most common biofuels used these days. Biofuels are expected to reduce the dependency on fossil fuels and thereby create a more sustainable and secure future for energy consumption. In addition, biofuels cause less carbon emission than fossil fuels and therefore, are less harmful to the

environment. In fact, research studies show that vehicles that use biofuels are more environmental-friendly [33].

Different biomass sources such as animal and organic municipal wastes, agricultural crops and residues, forest residues as well as some dedicated energy crops can be used to produce biofuels. Biofuels are generally classified in three categories based on their input biomass type. The most common biofuel category, known as the "first-generation", is produced from edible crops, which are also used as human food and/or animal feed. Currently, most of the biofuel produced in the world is first-generation. The U.S. is the largest ethanol producer in the world and most of its demand is satisfied domestically produced ethanol. However, the U.S. ethanol production using corn might have negative implications both for food production and prices as mentioned in Rajagopal et al. [42] and McNew and Griffith [36]. Similar concerns are also raised for Brazil that is the second largest ethanol producer in the world, on its ethanol production from sugarcane. Another example is China that started producing ethanol from corn in early 2000. This followed by a shortage of food and a drastic increase in food prices in China and therefore, the Chinese government banned use of grains for biofuel production in 2007 [49].

As Banerjee et al. [5] say, feedstocks that can be used for food and feed have been extensively utilized for biofuel production. This trend has shifted due to the concerns in food prices and focus is turned on to utilizing lignocellulosic biomass and non-edible biomass for "second-generation" or "advanced" biofuel production to enhance food security as well as energy security. This class of biofuel is produced from non-edible sources such as crop and forest residues, and dedicated energy crops like switchgrass. One of the most common and widely used second generation biofuels is cellulosic ethanol, known as c-ethanol [50].

One of the major sources for second-generation biofuels is forest biomass, also

known as woody biomass. Forest biomass includes trees and woody plants, and other woody parts grown in a forest, woodland, or rangeland environment, that are the by-products and/or residues of forest management. Products and/or by-products from production facilities such as sawmills, pulp and paper mills can also be refereed as woody biomass. Shabani et al. [47] argue that using forest biomass in bio-energy production has the potential to recover the waste that would otherwise be disposed to landfills or be incinerated, create jobs and also provide local and sustainable energy. Moreover, compared to other renewable energy sources such as solar and wind, the major advantage of using forest biomass is that it can be stored and used on demand [15]. Another biomass source for second-generation biofuel production is specially produced agricultural plants such as switchgrass and miscanthus. Switchgrass is a perennial energy crop native to North America. A well managed and maintained switchgrass for biofuel production can reach up to 10 feet with a life cycle of 10-20 years [22]. Switchgrass is well-adapted to grow in a large portion in the U.S. with low fertilizer applications and high resistance to naturally occurring pests and diseases. Moreover, as a perennial plant, switchgrass needs only to be planted once in long periods of time (more than 10 years) but it can be harvested annually using conventional equipment [28]. These advantages make switchgrass an important biomass alternative to produce advanced biofuels.

The third-generation biofuels are produced from algae based on a technology that is still in the development stage [13]. Nevertheless, the studies are promising and algaes will likely to increase the biomass portfolio to produce biofuels in the future.

Countries with high energy demands like the U.S. realized the trend in energy consumption and the concerns in environmental sustainability. Therefore, in 2005, the U.S. established its first renewable fuel volume mandate, which is called Renewable Fuel Standard (RFS) program and was created under the Energy Policy

Act (EPAct). 7.5 billion gallons of renewable fuel, i.e., biofuel, was required to be blended into gasoline by 2012 according to this mandate. In 2007, this program was extended and the amount of renewable fuel to be blended into gasoline was set to 36 billion gallons by 2022 [65]. Moreover, to avoid the negative effects of biofuel production on food market and prices, it was further specified that 16 out of this 36 billion gallons of biofuel must be advanced (second-generation) biofuels. However, the current annual worldwide advanced biofuel production capacity is less than one billion gallons [73]. This means that bio-energy investment, especially in advanced biofuel production, will substantially increase in coming years in the U.S. to satisfy these requirements. This leads us to one of the main motivations for this research.

1.2 Motivation

The recent trends in energy consumption, the urgent need for renewable energy alternatives for sustainable future and the energy programs like RFS, indicate a significant expansion in the biofuel industry in coming years. In order for this expansion to be done in a systematic way, the biomass-biofuel supply chain must be designed effectively. In other words, as for all production/distribution systems, the biofuel industry also needs a strong supply chain system to operate cost-effectively. This supply chain system starts from supply points that are farms, sawmills, landfills etc., and integrates with the oil supply chain at oil-refineries for blending after biofuel is produced. As An et al. [1] state that a comprehensive study on biomass-biofuel supply chain is required in order for the biofuel industry to compete with the long-standing oil industry, which has already a strong and efficient supply chain system. Therefore, this research focuses on creating tools to design the network structure of this expending supply chain system by determining optimal policies to reduce its cost and increase its profitability. These optimal policies include strategic decisions such

as facilities locations and biomass/biofuel pricing, which have effects on the system in the long run; as well as tactical and operational decisions such as transportation, production and inventory levels, which are taken periodically.

While designing its supply chain, important aspects of biomass-biofuel logistics must be addressed. One of these important aspects is the importance of designing an integrated biomass-biofuel supply chain system. Although some studies such as Eksioglu et al. [18], Huang et al. [26], Zhang et al. [74], address this issue, a more comprehensive study and modeling-solution framework, which is capable of handling large-scale problems, is needed to address the challenges in biomass-biofuel supply chain in an integrated fashion. Therefore, the study presented in Section 2 displays an integrated bio-energy supply chain system and modeling framework.

Another important aspect in this context is the uncertainty in biomass supply. Including uncertainty into the decision process makes the decision robust and mitigates the effect of variations. Hence, it is essential to incorporate uncertainty in the systems that are affected by variations. In most of the traditional supply chain systems the uncertainty is assumed to be on the demand side. However, in biomass-biofuel supply chain systems, the system uncertainty typically occurs on the supply side. Therefore, incorporating the effects of uncertainty in biomass supply due to environmental conditions, especially for crop related biomass, is highly important. We address the biomass supply uncertainty in the studies presented in Section 3 and Section 4.

Increasing the volume of dedicated energy crop production to produce more advanced biofuels is another important point in biomass-biofuel supply chain since the percentage of advanced biofuels will increase in overall biofuel production. However, studies such as Jensen et al. [28] and Villamil et al. [71] indicate that most farmers do not have enough knowledge about energy crops to adapt them. Moreover, most of

the farmers that know about the energy crops do not want to adopt them because of the uncertainty about their financial viability. For these reasons, policies involving economic incentives should be developed and implemented to increase the awareness on energy crops and to encourage farmers for more production. Therefore, we propose a policy involving a wholesale-price type of agreement in the study presented in Section 3.

Another motivation for this research is to investigate the relationship between price and supply chain network structure. Specifically, in biomass-biofuel supply chain context, we observe the effects of biomass price on the bio-energy supply chain network design in conjunction with the policy proposed in Section 3.

1.3 Dissertation Organization

The rest of this dissertation is organized as follows: The first study, which presents an integrated and deterministic bio-energy supply chain network design problem is described in Section 2. We introduce a biomass logistics network design problem, which considers a price-based biomass supply under uncertainty in Section 3. This study is extended to the whole bio-energy supply chain considering biofuel logistics as well as multiple-period planning horizon in Section 4. Lastly in Section 5, we summarize this dissertation, highlight its expected contributions and discuss future research directions.

Sections 2,3, and 4 are self-contained and correspond to a different study under the broad context of biomass-biofuel supply chain design. Hence, each of these three sections has its own introduction, problem definition and conclusion sections.

2. INTEGRATED BIO-ENERGY SUPPLY CHAIN NETWORK PLANNING PROBLEM

2.1 Introduction

The need for cleaner and more sustainable energy has become significant. One of the main reasons for the concern is our current dependence on fossil fuels, which are not sustainable and cause environmental pollution. Therefore, new and alternative ways to generate energy have become one of the top priorities. Biofuels, which are renewable sources of clean energy, are considered as one of future energy sources that can reduce the dependency on fossil fuels. Biofuels are less harmful to the environment than fossil fuels since they cause less carbon emission. Research studies show that these fuels decrease vehicle carbon emissions thus making them environmentally friendly [33]. Unlike fossil fuels, biofuels are renewable, hence, they are more sustainable. In addition, biofuels can provide new economic and political opportunities especially in developing countries [16]. Research is being conducted at an increasing rate on efficient production biofuels due to these benefits they provide.

Biofuels are produced from biomass, an organic product and/or waste. Different sources such as animal and organic municipal wastes, agricultural residues, forest residues as well as some dedicated energy crops such as switchgrass can be used as biomass. Biofuels are generally classified into three categories in terms of their input biomass type and the processing technology. The most common category biofuels, known as the first-generation, are produced from edible crops (i.e. sugar, corn etc.) which can also be used as animal and human food. Hence, some argue that production of first-generation biofuels might increase prices in the food industry and may lead to a global food crisis [44]. For these reasons, the development of second-

generation biofuels has subsequently received increased attention and much effort. This class of biofuel is produced from non-edible sources such as crop and forest residues, and dedicated energy crops like switchgrass. One of the most common and widely used second generation biofuels is cellulosic ethanol, known as c-ethanol [50]. The third-generation biofuels are produced from algae based on a technology that is largely still in the development stage [13]. Furthermore, obtaining biocrude using sludge from wastewater treatment is another method to produce biofuel [37].

As in every production/distribution system, the bio-energy industry also requires a sound design for its supply chain in order to operate effectively and efficiently. This supply chain system starts from the biomass supply points and integrates with the oil supply chain at the oil-refineries for blending after biofuel is produced. Therefore, an efficient supply chain system design is required for biomass as well as biofuel. An integrated biomass-biofuel supply chain logistics network design allows an effective overall system for improved efficiency. An et al. [1] argues that in order to compete with the already existing oil industry, which has a strong supply chain and logistic system, a comprehensive study on biomass-biofuel supply chain is needed.

In this study, our objective is multi-faceted including modeling, methodology, and application. *On the specific problem of interest and modeling side*, we consider the design of the whole supply chain including both the biomass and the biofuel logistics in an integrated fashion while simultaneously integrating strategic and tactical decisions pertaining to location, production, inventory, and distribution in a multi-period planning horizon setting. The decisions regarding the collection and biorefineries include both the location and capacity levels as well as inventory carried from period to period. As we explore in detail in Section 2.3, there are other studies addressing the supply chain design for bio-energy, however, not as comprehensively as the one provided herein with different aspects considered simultaneously. Most of the

studies consider three level supply chains and focus on its biomass side. Others consider again three level supply chain with both biomass and biofuel sides but do not take into account multiple biomass types, capacities and/or inventory decisions in a multi-period decision-making. Moreover, as opposed to previous studies, by allowing direct shipments from supply points to biorefineries, we also incorporate economies-of-scale in transportation which facilitates the use of collection centers by capturing the cost trade-off and/or transportation mode selection considerations.

On the methodology side, we observe that all of the studies in the related literature rely on a commercial software and solve small size instances for only case study purposes. In this section, we present an efficient Benders Decomposition based solution algorithm since state-of-the-art implementations of Branch-and-Cut approach is not able to handle realistically large-sized instances for design and analysis purposes. Although this implementation has been suggested by Rubin [45], we are not aware of specific applications that it has been successfully employed. In our bio-energy application, for the method to provide an acceptable performance, we develop valid constraints (surrogate constraints) which help us to obtain strong lower bounds and facilitate the generation of optimality cuts (rather than weaker feasibility cuts) in the Benders Decomposition framework. Although we develop these cuts in our context, they can be utilized in other generalized supply chain design settings that incorporate inventory and capacity decisions in a multi-period planning. We demonstrate the efficiency of our approach in experiments by solving many instances, which are organized in problem classes of varying sizes, with small known optimality gaps. We also note that efficient algorithms for generalized problems as in ours are also useful in solving stochastic versions of the problems as, most of the time, the latter relies on efficiently solving deterministic versions. For example, Chen and Fan [10] introduce a stochastic program in a three level network setting. However, in the context of

solution approach which is based on Lagrangean relaxation, the method relies on solving many deterministic problems (one for each scenario) efficiently.

Finally, *on the application and analysis side*, to illustrate the potential use of the suggested model and the solution methodology for policy decisions in realistic settings, we present a large case study utilizing realistic data from the entire state of Texas. We demonstrate the model's capabilities through a detailed analysis over changes in various model parameters. We also conduct a comprehensive analysis on the changes in network designs with respect to variations in a wide variety of important input parameters. For example, rather than quantifying the changes in the objective value with respect to changes in unit costs, we specifically study the network configuration impacts of geographically diverse biomass types, seasonality in biomass supplies, transportation economies-of-scale, supply and demand variability along with changes in conversion rates, transportation and production costs.

This section is organized as follows: In Section 2.2, we first provide a description of our problem and later present the notation and its mathematical model. This is followed by a discussion of the related literature that positions our study in context in Section 2.3. Section 2.4 is devoted to the solution algorithm which is based on Benders Decomposition (BD). A computational study that shows the efficiency of the solution algorithm is presented in Section 2.5. We present data gathering and results for our case study based on realistic data from Texas in Sections 2.6 and 2.7, respectively. Lastly, in Section 2.8, we provide our concluding remarks.

2.2 Problem Definition and Formulation

In the design of an integrated bio-energy supply chain network, we consider four levels (echelons) of facilities that comprise the overall system as depicted in Figure 2.1. They are (i) biomass supply points (farms, sawmills etc.) where the supply

(biomass) is produced, (ii) collection facilities where biomass is stored, (iii) biorefineries in which biomass is converted into biofuel, i.e., ethanol, and (iv) blending facilities, i.e., oil-refineries, where demand takes place. The objective is to minimize strategic and tactical logistics costs associated with this supply chain.

Biomass available at supply points, whose locations and yields are assumed to be known, is transported to collection facilities. Collection facilities are places where biomass is gathered and stored, but not mixed with each other. At supply points, the supply is time-varying due to environmental, climatic and economic factors. This variation in supply motivates the use of collection facilities. In periods where the supply is high, biomass can be stored in these facilities awaiting demand. However, biomass degrades over time. Hence, its value, i.e., usable amount needed to produce biofuel, decreases when kept in inventory. Another benefit of collection facilities is that they facilitate transportation economies-of-scale. Biomass transportation cost decreases when the transportation is done in large quantities. Therefore, unit transportation cost from collection facilities is typically less expensive than the unit transportation cost from supply points.

Biomass from supply points is transported to biorefineries either through collection facilities or directly for processing and conversion into biofuel. Unlike collection facilities, biorefineries are much bigger and complex structures, and hence, they have more expensive building and operating costs. Moreover, although their biomass storage capacities are less than those of collection facilities, biorefineries can store both biomass and biofuel.

From the biorefineries, biofuel is transported to blending facilities. Blending facilities are existing oil-refineries where gasoline is blended with biofuel. We assume that there is a certain number of blending facilities whose locations and demand for biofuel are known. According to this information, the problem is to determine:

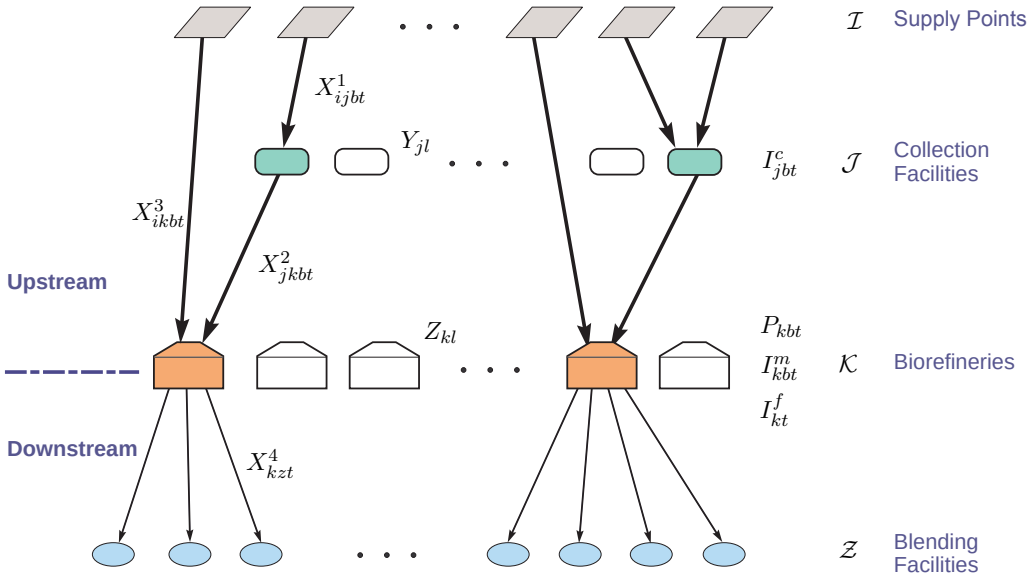


Figure 2.1: Bio-energy supply chain structure and notation

- the locations and the capacities of collection facilities and biorefineries to open;

and, for each planning period,

- shipment quantities of biomass from supply points to collection facilities and to biorefineries, and from collection facilities to biorefineries;
- shipment quantities of biofuel from biorefineries to blending facilities;
- production quantities of biofuel at each biorefinery; and
- inventory levels of biomass and biofuel at the collection facilities and biorefineries

while minimizing all the transportation and inventory holding costs, biomass processing costs, and the fixed costs associated with collection facilities and biorefineries.

To formulate our problem of interest, we first introduce the notation for sets, input parameters, and the decision variables as follows:

Sets:

$\mathcal{I}$	set of supply points, $i \in \mathcal{I}$
$\mathcal{J}$	set of collection facilities, $j \in \mathcal{J}$
$\mathcal{K}$	set of biorefineries, $k \in \mathcal{K}$
$\mathcal{Z}$	set of blending facilities, $z \in \mathcal{Z}$
$\mathcal{L}$	set of capacity (size) levels, $l \in \mathcal{L}$
$\mathcal{T}$	set of time periods, $t \in \mathcal{T}$
$\mathcal{B}$	set of biomass types, $b \in \mathcal{B}$

Parameters:

d_{uv}	distance between locations u and v $(u, v \in \mathcal{I} \cup \mathcal{J} \cup \mathcal{K} \cup \mathcal{Z})$
S_{ibt}	amount of biomass type b at supply point i at time t
D_{zt}	amount of biofuel demanded by blending facility z at time t
K_l	biofuel production capacity of a biorefinery with size l
β_b	conversion rate of one unit of biomass type b to one unit of biofuel
α_b	deterioration rate of biomass type b in one period
ω_b	processing cost of one unit of biomass type b
δ	discount factor for transportation cost between collection facilities and biorefineries
C_l^c	available inventory space for biomass of a collection facility with size l
C_l^m	available inventory space for biomass of a biorefinery with size l
C_l^f	available inventory space for biofuel of a biorefinery with size l
h_b^m	holding cost of one unit of biomass type b for one unit of time
h^f	holding cost of one unit of biofuel for one unit of time
r_b^m	per unit per mile biomass type b transportation cost
r^f	per unit per mile biofuel transportation cost

f^c_{jl} amortized fixed cost of opening and operating a collection facility of size l at location j

f^b_{kl} amortized fixed cost of opening and operating a biorefinery of size l at location k

Decision Variables:

X^1_{ijbt} amount of biomass type b shipped from supply point i to collection facility j at time t

X^2_{jkbt} amount of biomass type b shipped from collection facility j to biorefinery k at time t

X^3_{ikbt} amount of biomass type b shipped from supply point i to biorefinery k at time t

X^4_{kzt} amount of biofuel shipped from biorefinery k to blending facility z at time t

I^c_{jbt} amount of biomass type b kept in inventory at collection facility j at time t

I^m_{kbt} amount of biomass type b kept in inventory at biorefinery k at time t

I^f_{kt} amount of biofuel kept in inventory at biorefinery k at time t

Z_{kl} 1 if biorefinery of size l at k is open, 0 otherwise

Y_{jl} 1 if collection facility of size l at j is open, 0 otherwise

P_{kbt} amount of biomass type b processed at biorefinery k at time t.

According to the notation above, the problem is formulated as follows:

$$\begin{aligned}\text{Min}\quad & \sum_{t\in\mathcal{T}}\Bigg[\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{b\in\mathcal{B}} r_b^m\, d_{ij}\, X^1_{ijbt} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{b\in\mathcal{B}}(1-\delta)\, r_b^m\, d_{jk}\, X^2_{jkbt} \\ & + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{b\in\mathcal{B}} r_b^m\, d_{ik}\, X^3_{ikbt} + \sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{Z}} r^f\, d_{kz}\, X^4_{kzt}\Bigg] \\ & + \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} f^c_{jl}\, Y_{jl} + \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} f^b_{kl}\, Z_{kl} + \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}}\sum_{b\in\mathcal{B}} \omega_b\, P_{kbt} \\ & + \sum_{t\in\mathcal{T}}\Bigg[\sum_{j\in\mathcal{J}}\sum_{b\in\mathcal{B}} h_b^m\, I^c_{jbt} + \sum_{k\in\mathcal{K}}\sum_{b\in\mathcal{B}} h_b^m\, I^m_{kbt} + \sum_{k\in\mathcal{K}} h^f\, I^f_{kt}\Bigg] \qquad (2.1)\end{aligned}$$

subject to

$$\sum_{b\in\mathcal{B}} \beta_b\, P_{kbt} \le \sum_{l\in\mathcal{L}} K_l\, Z_{kl} \qquad \forall\, k\in\mathcal{K},\, t\in\mathcal{T} \qquad (2.2)$$

$$\sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}} X^1_{ijbt} \le \sum_{l\in\mathcal{L}} C^c_l\, Y_{jl} \qquad \forall\, j\in\mathcal{J},\, t\in\mathcal{T} \qquad (2.3)$$

$$\sum_{b\in\mathcal{B}} I^c_{jbt} \le \sum_{l\in\mathcal{L}} C^c_l\, Y_{jl} \qquad \forall\, j\in\mathcal{J},\, t\in\mathcal{T} \qquad (2.4)$$

$$\sum_{b\in\mathcal{B}} I^m_{kbt} \le \sum_{l\in\mathcal{L}} C^m_l\, Z_{kl} \qquad \forall\, k\in\mathcal{K},\, t\in\mathcal{T} \qquad (2.5)$$

$$I^f_{kt} \le \sum_{l\in\mathcal{L}} C^f_l\, Z_{kl} \qquad \forall\, k\in\mathcal{K},\, t\in\mathcal{T} \qquad (2.6)$$

$$\sum_{j\in\mathcal{J}} X^1_{ijbt} + \sum_{k\in\mathcal{K}} X^3_{ikbt} \leq S_{ibt} \qquad \forall\, i\in\mathcal{I},\, b\in\mathcal{B},\, t\in\mathcal{T} \quad (2.7)$$

$$\sum_{i\in\mathcal{I}} X^1_{ijbt} + (1-\alpha_b)\, I^c_{jb(t-1)} = \sum_{k\in\mathcal{K}} X^2_{jkbt} + I^c_{jbt} \qquad \forall\, j\in\mathcal{J},\, b\in\mathcal{B},\, t\in\mathcal{T} \quad (2.8)$$

$$\sum_{i\in\mathcal{I}} X^3_{ikbt} + \sum_{j\in\mathcal{J}} X^2_{jkbt} + (1-\alpha_b)\, I^m_{kb(t-1)} = P_{kbt} + I^m_{kbt} \qquad \forall\, k\in\mathcal{K},\, b\in\mathcal{B},\, t\in\mathcal{T} \quad (2.9)$$

$$\sum_{b\in\mathcal{B}} \beta_b\, P_{kbt} + I^f_{k(t-1)} = \sum_{z\in\mathcal{Z}} X^4_{kzt} + I^f_{kt} \qquad \forall\, k\in\mathcal{K},\, t\in\mathcal{T} \quad (2.10)$$

$$\sum_{l\in\mathcal{L}} Z_{kl} \leq 1 \qquad \forall\, k\in\mathcal{K} \quad (2.11)$$

$$\sum_{l\in\mathcal{L}} Y_{jl} \leq 1 \qquad \forall\, j\in\mathcal{J} \quad (2.12)$$

$$\sum_{k\in\mathcal{K}} X^4_{kzt} \geq D_{zt} \qquad \forall\, z\in\mathcal{Z},\, t\in\mathcal{T} \quad (2.13)$$

$$X^1_{ijbt},\, X^2_{jkbt},\, X^3_{ikbt},\, X^4_{kzt},\, I^c_{jbt},\, I^m_{kbt},\, I^f_{kt},\, P_{kbt} \geq 0 \qquad \forall\, i,\, j,\, k,\, z,\, b,\, t \quad (2.14)$$

$$Z_{kl},\, Y_{jl} \in \{0,1\} \qquad \forall\, j,\, k,\, l. \quad (2.15)$$

In the objective function (2.1), the first group denotes the transportation costs from supply points to collection facilities, collection facilities to biorefineries, supply points to biorefineries and biorefineries to blending facilities, respectively. The second and third terms cover fixed cost associated with opening and operating collection facilities and biorefineries. Fourth term covers the processing cost of biomass. The last group includes biomass inventory holding costs at collection facilities, and inventory holding costs of biomass and biofuel at biorefineries.

Constraint (2.2) represents the production (process) capacity at biorefineries. This constraint also sets the production to zero if the biorefinery is not open. Constraint (2.3) ensures that biomass is transported to a collection facility only when it is open while, at the same time, dictating the capacity limitation for an open collection facility. Constraints (2.4)-(2.6) are inventory capacity constraints at collection facilities and biorefineries. Constraints (2.7) ensure that biomass outflow from a supply point is less than its supply availability. Flow balance constraints for collection facilities and biorefineries are addressed by constraints (2.8)-(2.10), respectively, which ensure the equality of incoming and outgoing biomass and biofuel flows at corresponding facilities. Constraints (2.11) and (2.12) make sure that no more than one facility (for biorefineries and collection facilities, respectively) can be open at a given location. Demand satisfaction is ensured in constraint (2.13). Lastly, constraints (2.14) and (2.15) define nonnegativity and integrality requirements for the decision variables, respectively.

2.3 Related Literature

The literature on the biomass-biofuel supply chain is developing at an increasing pace recently. Several studies with varying focus on its different aspects appear in a variety of journals. Below, we present a review of the related literature from a wide

range of outlets and highlight our contributions in the logistical modeling, solution method, and analysis contexts.

First, we mention three review papers that appeared in the literature recently. An et al. [2] provide a review on biofuel and petroleum supply chains. In that study, the authors categorize the related studies according to their focus areas as upstream (biomass) and downstream (biofuel) while specifically noting that the most of the studies in the field focus on the upstream supply chain. Sharma et al. [48] provide a comprehensive review on biomass-biofuel supply chain designs and discuss the energy needs and targets, feedstocks and conversion processes. A classification based on several aspects including decision levels, supply chain structure, and modeling approaches is presented. Among other observations such as the common focus on cost considerations, quantitative modeling and the use of realistic data, authors also conclude main focus in the context of facilities location and network design is on the upstream supply chain rather than the design of an integrated chain. In addition, Shabani et al. [47] provide a literature review on the supply chains related to electricity, heat and biofuels production from forest biomass and provide extensive discussion on the deterministic and stochastic mathematical models including their assumptions and decision variables in the context of forest biomass supply chains.

Studies that are most related to our work are found in the area focusing on strategic and tactical decisions in biofuel supply chain management as summarized next.

The setting that is most closely related to ours is studied by Eksioglu et al. [18] which considers both the upstream and downstream decisions as well as material flows. Specifically, a four level supply chain system, including biomass supply, collection facilities, biorefineries, and blending facilities, is modeled to determine collection facility locations and biorefinery location and capacity decisions. The model

has a multi-period setting and includes harvesting, inventory, transportation and processing costs. The authors conduct a case study in Mississippi where they consider 45 counties as supply points. Other data related to problem size (such as the number of potential and opened collection/biorefinery facility locations, variations in supply and demand values from period to period, etc.) are not reported explicitly. Analysis indicates that increasing biorefinery construction, unit transportation, processing and biomass collection costs increase the unit final delivery cost (equivalently, the total cost of satisfying the total fixed demand) of c-ethanol. Also, a decreased conversion rate (assumed to indicate less biomass availability) increases transportation distances and, in turn, costs due to a need for larger collection area for higher biomass volume. In our analysis, we consider more parameters, that are beyond increasing only unit input costs, and their impact on the overall network structure, rather than the impact on the total cost of delivery which is expected to increase by increasing unit operational costs under fixed demand. From the modeling perspective, although our model shows some similarities with the one by Eksioglu et al. [18], there are some major differences. One of the major differences is the demand satisfaction constraint. In the study by Eksioglu et al. [18], although demand is defined specific for each location and period, in the model construction, only an aggregate demand for a period is satisfied rather than satisfying the demand for each location. In our model, we ensure that demand is satisfied for each blending facility and for each time period (by our constraint (2.13)). Secondly, as opposed to our model, in their model, Ekşioğlu et al. do not consider any fixed costs associated with collection facilities (f^c_{jl} in our model). Thus, these facilities can be opened freely anywhere in the network without any cost trade-off implications. Furthermore, in Eksioglu et al. [18], capacity limitations at collection facilities are not considered, as opposed to our model which includes constraint (2.3) for this purpose. In fact, these terms on

collection facilities are necessary to incorporate their location decisions correctly into the model. In terms of the solution methodology, as is the case with all the other studies in the area, Eksioglu et al. [18] utilize an off-the-shelf software (CPLEX) to solve relatively small size problems. Although exact problem sizes are not provided in their study, a comment on future research is made regarding the need to develop efficient solution procedures. In this study, we provide an efficient Benders Decomposition algorithm to solve our model which is more general and challenging than the model provided by Eksioglu et al. [18]. We consider a specific implementation and also develop valid constraints to be used as surrogate for improved algorithmic efficiency.

Huang et al. [26] develop an optimization model for a three level supply chain (supply points, biorefineries, and demand points) in a multi-period context that determines the biorefinery locations and the material flow to minimize the total cost including procurement, transportation, production, and location costs. The model does not allow for any collection facilities. Also, no end-of-period inventory holding is allowed, but shortage at demand sites is penalized. Authors conduct a case study in California in which 8 biomass types with a number of supply points ranging between 14 to 57, 28 candidate biorefinery sites and 143 demand locations (cities) are considered and the resulting model is solved using off-the-shelf software (CPLEX). Moreover, three input parameters, transportation cost, maximum refinery capacity and feedstock availability, are altered and their effects on the system cost and design are found to be insignificant with the exception that the decreased refinery capacities lead to changes in system design, specifically leading to higher number of refineries.

Zhu and Yao [75] consider the upstream biomass chain as three levels (supply, collection, and biorefinery points). They present a model to determine the collection and biorefinery locations along with production, inventory and transportation

decisions for each time period with the objective of maximizing the system revenue. Capacity limitations for production at the biorefinery and for storage at collection facilities are also considered. In their case study, the authors consider 12 periods (months) and an annual planning horizon along with 3 biomass types with 14 (2, 2, and 10) supply locations, 3 candidate collection sites and 2 candidate biorefinery locations. The model is solved using off-the-shelf software (CPLEX). Their analysis focuses on the differences between having only one type of biomass (switchgrass) vs all three types in determining facility locations and biomass flows. It is concluded that, in the latter case, the total supply can increase and help to smooth the biofuel production by leveling fluctuations due to seasonality in switchgrass, and in turn, increase the unit profit from biofuel. Zhu et al. [76] consider a similar model with the same setting and case data but with only one biomass type and no purchasing costs.

An et al. [1] present a supply chain model with five levels including supply, preprocessing, biorefinery, warehouse and demand locations. With a profit maximization objective, the model determines the locations of preprocessing, biorefinery and warehouse locations and capacities (as continuous variables) and transportation arcs with associated fixed and variable costs. In their case study, nine locations (counties) in Central Texas are considered as the supply, demand and candidate locations along with a one-year planning horizon with four periods. The model is solved using off-the-shelf software (CPLEX) and the effects of changing biomass cost and yield, and biofuel price and demand on the supply chain are summarized.

Zhang et al. [74] also provide a model for four level supply chain by considering switchgrass as the biomass type to determine switchgrass harvesting method as well as biorefinery locations and capacity levels and preprocessing facilities locations. The objective is to minimize total system cost which includes harvesting, storage,

transportation, processing and fix costs. A case study is conducted using all 53 counties in North Dakota to demonstrate the capabilities of the model. One year planning horizon with each time period representing a month is considered. The authors use off-the-shelf software (LINGO 10.0) to solve the problem. The results of the case study show that loose chop is the optimal harvesting method for switchgrass as compared to traditional bailing methods. The authors also claim that biorefinery locations are insensitive to the annual variation in switchgrass yield which is obtained from only 28 counties and each of the 28 counties include a preprocessing facility.

Bowling et al. [9] present a model for a three level upstream network that aims to maximize the total system profit while determining the preprocessing and processing facility locations as well as the flow between these facilities in only a single period setting and a single biomass type. The authors conduct two small case studies with 6 supply locations and 2 candidate locations for preprocessing and processing facilities and solve their model using GAMS. The results show that dispersed facility locations usually yield better solutions.

More recently, Chen and Onal [11] propose a mixed integer nonlinear programming model that incorporates feedstock price, harvesting decisions, and biorefinery locations for profit maximization and develop an iterative heuristic procedure for its solution. The model determines the response in biomass feedstock supply in the U.S. to meet the biofuel production targets considering biorefinery location decisions and conclude that increased biofuel demand imposed by government mandates would lead to significant increase in crop prices and the number of biorefineries. In a similar study, by considering existing biorefineries explicitly, Marvin et al. [35] suggest model, with embedded cash flow analysis, for a three level biofuel supply chain that determines biomass processing facility locations and capacities along with biomass processing technologies. The authors conduct a case study in Midwest U.S. and

argue the feasibility of the renewable fuel standard mandates for 2015.

Some recent studies also consider uncertainty and develop stochastic programs. Kim et al. [29] present a profit maximization model for determining locations of two level conversion facilities, with given supply and demand points, under uncertainty. After setting five parameters as the most influential on the objective function, a two stage stochastic program with 32 scenarios (i.e., 2^5 in total with high and low values for each parameter) is solved for data from Southeast U.S. with 30 supply, 29 (level 1) and 10 (level 2) candidate conversion facilities and 10 demand locations using Monte Carlo simulation in which the model with any randomly generated input is solved via CPLEX. Chen and Fan [10] introduce a two-stage stochastic program to model bioethanol supply chain in which the refinery and terminal locations are determined in the first-stage and expected cost of transportation and bioethanol production is considered in the second-stage. The authors utilize progressive hedging approach after introducing variable splitting and relaxing the copy constraints for Lagrangean subproblem decomposition for each scenario. Each such subproblem is essentially a full-scale deterministic problem and solved using CPLEX. A case study in California, using the same data as in Huang et al. [26], is conducted to explore the potential of waste-based bioethanol production. Only two settings are considered where demand and supply uncertainties are addressed with 4 and 10 scenarios. Studies that consider uncertainty will be discussed more in detail in Section 3.2

We observe that computational difficulties still persist, especially more so with the stochastic programs as highly limited number of scenarios (likely not nearly adequate to represent the underlying nature of the uncertainty) are used to obtain results. This fact also highlights the need for improved solution approaches such as the one presented in this section to solve deterministic large size instances so that they can be utilized in obtaining solutions for stochastic counterparts.

2.4 Solution Methodology - Benders Decomposition (BD)

BD algorithm decomposes the overall MIP problem into a master problem and a subproblem [6]. The master problem (BMP) includes all the integer variables in the original MIP and the associated constraints. A subproblem (BSP) is obtained by assuming a fixed set of values for all integer variables in the original MIP, and, thus, it is a linear programming problem that includes all the continuous decision variables and their associated constraints along the terms in the objective function of the original MIP. Its dual (DBSP) is easily obtained and utilized to generate a Benders cut that is later added to the BMP via the use of an auxiliary continuous variable that represents a lower bound on the subproblem portion of the original problem. In an iterative scheme, the solution of a BMP provides a set of integer variable values along with a lower bound on the overall (minimization) problem. This solution is used to construct a DBSP whose solution provides a set of dual variable values to be used to generate a Benders cut. The solution to BDSP also facilitates the calculation of an upper bound as it corresponds to generating a feasible solution to the overall problem with fixed integer variable values. Cut generation via DBSP solution and cut addition to BMP and construction of new DBSP via BMP solution steps are continued in successive iterations until an acceptable gap between upper and lower bounds is achieved. BD algorithm is guaranteed to converge to an optimal solution since the BMP essentially corresponds to a relaxation of the original MIP in which some Benders cuts are added as in delayed constraint generation; If all possible Benders cuts were available and included in BMP, then the BMP is equivalent to the original MIP.

2.4.1 Benders Subproblem (BSP)

We obtain the BSP by assuming given values of integer variables $\hat{Z}_{kl}$ and $\hat{Y}_{jl}$ in our MIP problem (2.1)-(2.15). That is, in Figure 2.1, fixing the locations and the capacities of the collection facilities and biorefineries, we obtain our BSP which is a generalized network flow problem given as follows:

$$\begin{aligned}\text{Min} \quad & \sum_{t\in\mathcal{T}}\Bigg[\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{b\in\mathcal{B}} r_b^m\, d_{ij}\, X^1_{ijbt} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{b\in\mathcal{B}}(1-\delta)\, r_b^m\, d_{jk}\, X^2_{jkbt} \\ & + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{b\in\mathcal{B}} r_b^m\, d_{ik}\, X^3_{ikbt} + \sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{Z}} r^f\, d_{kz}\, X^4_{kzt}\Bigg] \\ & + \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}}\sum_{b\in\mathcal{B}} \omega_b\, P_{kbt} + \sum_{t\in\mathcal{T}}\Bigg[\sum_{j\in\mathcal{J}}\sum_{b\in\mathcal{B}} h_b^m\, I^c_{jbt} + \sum_{k\in\mathcal{K}}\sum_{b\in\mathcal{B}} h_b^m\, I^m_{kbt} + \sum_{k\in\mathcal{K}} h^f\, I^f_{kt}\Bigg]\end{aligned} \tag{2.16}$$

subject to $\quad (2.7), (2.8), (2.9), (2.10), (2.13), (2.14),$

$$\sum_{b\in\mathcal{B}} \beta_b\, P_{kbt} \le \sum_{l\in\mathcal{L}} K_l\, \hat{Z}_{kl} \qquad \forall\, k\in\mathcal{K},\, t\in\mathcal{T} \tag{2.17}$$

$$\sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}} X^1_{ijbt} \le \sum_{l\in\mathcal{L}} C^c_l\, \hat{Y}_{jl} \qquad \forall\, j\in\mathcal{J},\, t\in\mathcal{T} \tag{2.18}$$

$$\sum_{b\in\mathcal{B}} I^c_{jbt} \le \sum_{l\in\mathcal{L}} C^c_l\, \hat{Y}_{jl} \qquad \forall\, j\in\mathcal{J},\, t\in\mathcal{T} \tag{2.19}$$

$$\sum_{b\in\mathcal{B}} I^m_{kbt} \le \sum_{l\in\mathcal{L}} C^m_l\, \hat{Z}_{kl} \qquad \forall\, k\in\mathcal{K},\, t\in\mathcal{T} \tag{2.20}$$

$$I^f_{kt} \le \sum_{l\in\mathcal{L}} C^f_l\, \hat{Z}_{kl} \qquad \forall\, k\in\mathcal{K},\, t\in\mathcal{T}. \tag{2.21}$$

Defining ψ_{ibt}, θ_{jbt}, λ_{kbt}, μ_{kt}, ρ_{zt}, ζ_{kt}, η_{jt}, κ_{jt}, ξ_{kt}, and χ_{kt} as dual variables associated with constraints (2.7), (2.8), (2.9),(2.10), (2.13), (2.17), (2.18), (2.19), (2.20), and

(2.21), respectively, the dual Benders subproblem DBSP is constructed as follows:

$$\begin{aligned}\text{Max} \quad & \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} K_l \hat{Z}_{kl}\,\zeta_{kt} + \sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} C_l^c \hat{Y}_{jl}\,\eta_{jt} + \sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} C_l^c \hat{Y}_{jl}\kappa_{jt} \\ & + \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} C_l^m \hat{Z}_{kl}\,\xi_{kt} + \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} C_l^f \hat{Z}_{kl}\,\chi_{kt} + \sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}}\sum_{t\in\mathcal{T}} S_{ibt}\,\psi_{ibt} \\ & + \sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} D_{zt}\,\rho_{zt} \qquad (2.22)\end{aligned}$$

subject to

$$\eta_{jt} + \psi_{ibt} + \theta_{jbt} \le r_b^m\, d_{ij} \qquad \forall\, i\in\mathcal{I},\ j\in\mathcal{J},\ b\in\mathcal{B},\ t\in\mathcal{T} \qquad (2.23)$$

$$-\theta_{jbt} + \lambda_{kbt} \le (1-\delta)\, r_b^m\, d_{jk} \qquad \forall\, j\in\mathcal{J},\ k\in\mathcal{K},\ b\in\mathcal{B},\ t\in\mathcal{T} \qquad (2.24)$$

$$\psi_{ibt} + \lambda_{kbt} \le r_b^m\, d_{ik} \qquad \forall\, i\in\mathcal{I},\ k\in\mathcal{K},\ b\in\mathcal{B},\ t\in\mathcal{T} \qquad (2.25)$$

$$-\mu_{kt} + \rho_{zt} \le r^f\, d_{kz} \qquad \forall\, k\in\mathcal{K},\ z\in\mathcal{Z},\ t\in\mathcal{T} \qquad (2.26)$$

$$\kappa_{jt} + (1-\alpha_b)\,\theta_{jb(t+1)} - \theta_{jbt} \le h_b^m \qquad \forall\, j\in\mathcal{J},\ b\in\mathcal{B},\ t\in\mathcal{T} \qquad (2.27)$$

$$\xi_{kt} + (1-\alpha_b)\,\lambda_{kb(t+1)} - \lambda_{kbt} \le h_b^m \qquad \forall\, k\in\mathcal{K},\ b\in\mathcal{B},\ t\in\mathcal{T} \qquad (2.28)$$

$$\chi_{kt} + \mu_{k(t+1)} - \mu_{kt} \le h^f \qquad \forall\, k\in\mathcal{K},\ t\in\mathcal{T} \qquad (2.29)$$

$$\beta_b\,\zeta_{kt} - \lambda_{kbt} + \beta_b\,\mu_{kt} \le \omega_b \qquad \forall\, k\in\mathcal{K},\ b\in\mathcal{B},\ t\in\mathcal{T} \qquad (2.30)$$

$$\zeta_{kt},\ \eta_{jt},\ \kappa_{jt},\ \xi_{kt},\ \chi_{kt},\ \psi_{ibt} \le 0 \quad \rho_{zt} \ge 0 \qquad \forall\, i\in\mathcal{I},\ j\in\mathcal{J},\ k\in\mathcal{K},\ b\in\mathcal{B},\ t\in\mathcal{T} \qquad (2.31)$$

$$\theta_{jbt},\ \lambda_{kbt},\ \mu_{kt} \quad \text{unrest} \qquad \forall\, j\in\mathcal{J},\ k\in\mathcal{K},\ z\in\mathcal{Z},\ b\in\mathcal{B},\ t\in\mathcal{T}. \qquad (2.32)$$

Let $\mathcal{E}$ denote the set of all extreme points of the DBSP polyhedron given by (2.23)-

(2.32) and ζ^e_{kt}, η^e_{jt}, κ^e_{jt}, ξ^e_{kt}, χ^e_{kt}, ψ^e_{ibt} and ρ^e_{zt} and $Dvar^e$ denote the associated dual variable and objective function value with extreme point $e \in \mathcal{E}$. Furthermore, let $Dvar^*$ be the optimal value for the portion of original MIP's objective value employed in BSP. Then, since $Dvar^e \leq Dvar^*$, $\forall e \in \mathcal{E}$, BDSP can be restated as $\min_{Dvar \geq 0}\{Dvar : Dvar^e \leq Dvar, \ \forall e \in \mathcal{E}\}$ where

$$\begin{aligned} Dvar^e &= \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} K_l \hat{Z}_{kl}\, \zeta^e_{kt} + \sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} C^c_l \hat{Y}_{jl}\, \eta^e_{jt} + \sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} C^c_l \hat{Y}_{jl} \kappa^e_{jt} \\ &+ \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} C^m_l \hat{Z}_{kl}\, \xi^e_{kt} + \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} C^f_l \hat{Z}_{kl}\, \chi^e_{kt} + \sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}}\sum_{t\in\mathcal{T}} S_{ibt}\, \psi^e_{ibt} \\ &+ \sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} D_{zt}\, \rho^e_{zt}. \end{aligned} \tag{2.33}$$

2.4.2 Reformulation and the Benders Master Problem (BMP)

Using the above representation of DBSP, we write a reformulation of the original MIP as

$$\text{Min} \quad \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} f^c_{jl}\, Y_{jl} + \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} f^b_{kl}\, Z_{kl} + Dvar \tag{2.34}$$

subject to

(2.11), (2.12),

$$Dvar^e \leq Dvar \qquad \forall\, e \in \mathcal{E} \tag{2.35}$$

$$Z_{kl},\, Y_{jl} \in \{0,1\}, \qquad \forall\, j \in \mathcal{J},\, k \in \mathcal{K},\, l \in \mathcal{L}, \tag{2.36}$$

$$Dvar \geq 0. \tag{2.37}$$

Notice that constraint (2.35) is written for all the extreme points in *DBSP* polyhedron, i.e., set $\mathcal{E}$ which is not available. In the BD framework, these constraints are generated in a delayed fashion, one at a time in each iteration, since not all the constraints in (2.35) will be binding at optimality. Thus, the reformulation with only a subset of constraints denoted by $\mathcal{E}^S$, at any iteration of the algorithm. This relaxed problem, Benders master problem (BMP) whose optimal solution provides a lower bound on the original MIP, is the same as above reformulation in which the constraint (2.35) is only for $e \in \mathcal{E}^S$.

2.4.3 Surrogate Constraints for BMP

One of the issues that arise in ensuring the efficiency of the BD algorithm is related to the feasibility of the BSP (or the boundedness of the DBSP) which is dependent on the BMP solution that it receives as input. In our case, to generate a Benders cut for BMP, the subproblem receives the $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ values that are determined by the most recent BMP solution. For this set of $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ values, if the BSP is feasible, then an optimality Benders cut is generated in the form given by (2.35), for otherwise, a feasibility Benders cut is generated based on extreme rays of the (unbounded) DBSP polyhedron. The feasibility cut are typically lead to inefficiency of the BD algorithm since they do not help to improve the lower bound the BMP provides. To be able to always add optimality cuts, i.e., to ensure that the $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ values passed to a subproblem always lead to a feasible BSP, we devise surrogate constraints, valid but not necessary for the correctness of the original MIP, and keep them in the BMP. For this purpose, we develop three types of surrogate constraints.

The first constraint, (2.38), ensures that, for each period in the planning horizon, there is enough cumulative biorefinery production capacity available to produce the total cumulative demand up to the end of that period.

$$\sum_{z\in\mathcal{Z}}\sum_{v=1}^{t} D_{zv} \leq t \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} K_l\, Z_{kl} \qquad \forall\, t \in \mathcal{T}. \tag{2.38}$$

The left-hand side of (2.38) denotes the total demand for the first t time periods while its right-hand side is the total available biorefinery capacity (established in the beginning of the planning horizon) multiplied by the number of time periods elapsed, i.e., t.

Before stating the next two surrogate constraints, we define terms used in their construction. Let DSI_t denote the minimum amount of biofuel demand that must be satisfied from the inventory at period $t \in \mathcal{T}$, i.e., $DSI_t = \sum_{z\in\mathcal{Z}} D_{zt} - \sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}} \beta_b S_{ibt}$. Notice that if $DSI_t \leq 0$, the total demand in period t can be satisfied from the total available supply in period t. On the other hand, if $DSI_t > 0$, at least DSI_t equivalent of biofuel inventory (either as biofuel or as biomass) must be on hand in the system to satisfy demand in period t. Considering all sub-intervals of the planning horizon $\mathcal{T}$ and taking the one with the maximum sum of DSIs gives the amount of biofuel that must be satisfied from the inventory in the most extreme case. Let us denote this value as the biofuel-equivalent (i.e., convertible and/or usable as biofuel) Minimum Required Inventory Capacity in the system, $MRIC = \max_{v,t\in\mathcal{T},v\leq t}(\sum_{m=v}^{t} DSI_m)$. We know that the system biofuel-equivalent inventory capacity must be greater than $MRIC$. Moreover, let $SSPC_t$ denote the sum of surplus biofuel production capacity up to and including time period t, i.e., $SSPC_t = t\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} K_l Z_{kl} - \sum_{z\in\mathcal{Z}}\sum_{v=1}^{t} D_{zv}$. Note that constraint (2.38) ensures the nonnegativity of $SSPC_t$. With these definitions, we next provide the surrogate constraints that relate $MRIC$ to biofuel-equivalent capacities, (2.39) and (2.40).

$$MRIC \leq \max_b(\beta_b)\left(\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} C_l^m Z_{kl} + \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} C_l^c Y_{jl}\right) + \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} C_l^f Z_{kl} \quad (2.39)$$

$$MRIC \leq \max_b(\beta_b)\left(\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} C_l^m Z_{kl} + \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} C_l^c Y_{jl}\right) + SSPC_t \quad \forall\, t \in \mathcal{T}. \quad (2.40)$$

In (2.39), the first term on the right hand side represents the minimum necessary biofuel-equivalent biomass inventory capacity. This is ensured via the use of the largest biofuel conversion factor, $\max_b(\beta_b)$ while finding the biofuel-equivalent amount of total biomass capacity. The second term represents the total biofuel inventory capacity in the system. Therefore, constraint (2.39) ensures that enough biofuel-equivalent inventory capacity is made available at opened collection and biorefinery locations to surpass $MRIC$ in the system.

In (2.40), the first term on the right hand side is the same as that of (2.39). The second term, on the other hand, is the sum of surplus biofuel production capacity, $SSPC_t$, available at time period t. Therefore, constraint (2.40) guarantees that there are open facilities (collection and biorefinery locations) that can provide enough total biofuel-equivalent capacity to exceed $MRIC$.

2.4.4 BD Implementation

As an alternative implementation to base BD algorithm described above, we also implement the BD algorithm by using callbacks whenever a new Benders cuts can be generated in the master problem solution [45]. Callback function is provided by CPLEX 12.4 [27]. Specifically in this approach, called **BD-Callback** herein, BMP branch-and-cut solution tree is constructed and solved only once, unlike a BD implementation where BMP is resolved (and, thus, its solution tree is reconstructed) after addition of a new Benders cut in every iteration of the algorithm. In BD-

Callback implementation, whenever a feasible solution (incumbent solution) is found while solving *BMP*, *BDSP* is constructed using the corresponding $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ values and solved to generate a new Benders cut. This cut is then added to *BMP* as a "Lazy Constraint" and the process of solving *BMP* incorporating this constraint is resumed on the current solution tree. Hence, in BD-Callback implementation, outlined in Algorithm 1 while several *BDSP*s are solved in the overall process, only one instance of *BMP* is solved instead of solving *BMP* and *BDSP* exclusively in every iteration as in a base BD Algorithm.

Algorithm 1 BD-Callback Implementation

initialize $\epsilon = 0.02$, `optgap` $= 1.0$, `Runtime`$=0$, `Stoptime`$=7200$
Start solving **BMP**
while (`Runtime` $\leq$ `Stoptime` **and** `optgap` $> \epsilon$) **do**
  Continue solving **BMP**
  if A new incumbent solution is found with $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ **then**
    {*Callback Function*}
    Substitute $\hat{Y}_{jl}$ and $\hat{Z}_{kl}$ and solve **DBSP**
    Generate a Benders cut and add it to BMP as a Lazy Constraint
  end if
  Record `optgap` and `Runtime`
end while
return BMP objective value and its corresponding solution

We note that BD-Callback implementation is similar to **ϵ-optimal** implementation of BD algorithm [23, 70] in which the *BMP* is constructed with an additional bound constraint, given as follows for our problem:

$$\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} f^c_{jl}\, Y_{jl} + \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} f^b_{kl}\, Z_{kl} + Dvar \leq UB\,(1-\epsilon).$$

In every BD iteration, *BMP* is solved until only an integer feasible solution is obtained which has an objective value less than 100ϵ % of the current (best) UB. Clearly, if this fails, then the best UB is an ϵ-optimal solution. The disadvantage of the ϵ-optimal approach stems from the fact that the verification of infeasibility of *BMP* takes increasingly longer runtime as the iterations progress. In turn, the potential advantage of the BD-Callback implementation over the ϵ-optimal implementation is due not restarting the solution of *BMP* from scratch in each iteration of the algorithm. In our computational testing, we also considered an ϵ-optimal implementation, however we could not observe any significant improvement on the solution performance provided by a base BD Algorithm. However, since the BD-Callback implementation provided improved runtimes over the base BD algorithm for large instances, we employ the latter in our computational studies.

2.5 Computational Study on Algorithmic Performance

To test the effectiveness of the proposed solution algorithm based on comparisons to Branch-and-Cut (B&C) approach, we generate random data sets that represent varying values of input parameters. For B&C solutions, we employ CPLEX 12.4 with default settings for cut generation, preprocessing, and upper bound heuristics. All the experiments are conducted using C++ with STL (Standard Template Library) and Concert Technology (ILOG, Inc.) on machines with Intel Core2 3.00 GHz processor and 8 GB RAM running 64-bit OS.

2.5.1 Data Generation

To test the performance of the suggested BD-Callback approach, we employ 16 data classes, each with randomly generated 10 instances as shown in Table 2.1. Test classes are obtained by considering two values for each of the four parameters affecting the problem size as follows: The number of supply points and potential

collection facility location pairs, $|\mathcal{I}| - |\mathcal{J}| = 200 - 20, 300 - 30$; potential biorefinery locations, $|\mathcal{K}| = 20, 30$; blending facilities, $|\mathcal{Z}| = 25, 35$; and time periods $|\mathcal{T}| = 5, 10$. Moreover, three biomass types and two capacity sizes for both collection facilities and biorefineries are considered in all 16 data classes.

Table 2.1: Data classes and their sizes

Class	$\|\mathcal{I}\|$	$\|\mathcal{J}\|$	$\|\mathcal{K}\|$	$\|\mathcal{Z}\|$	$\|\mathcal{T}\|$	Class	$\|\mathcal{I}\|$	$\|\mathcal{J}\|$	$\|\mathcal{K}\|$	$\|\mathcal{Z}\|$	$\|\mathcal{T}\|$
C1	200	20	20	25	5	C9	200	20	20	25	10
C2	300	30	20	25	5	C10	300	30	20	25	10
C3	200	20	30	25	5	C11	200	20	30	25	10
C4	300	30	30	25	5	C12	300	30	30	25	10
C5	200	20	20	35	5	C13	200	20	20	35	10
C6	300	30	20	35	5	C14	300	30	20	35	10
C7	200	20	30	35	5	C15	200	20	30	35	10
C8	300	30	30	35	5	C16	300	30	30	35	10

A geographical area of 500x500 miles is considered and all location coordinates are randomly generated inside this area. Euclidean distances are then calculated between all the supply, demand and potential facility points. All the other parameters are randomly generated using uniform distribution as summarized in Table 2.2.

Table 2.2: Parameter distributions

Parameter	S_{ibt}	D_{zt}	β_b	α_b
Distribution	U[0, 10000]	U[5M, 6M]	U[60, 70]	U[0.05, 0.06]
Parameter	ω_b	h_b^m	r_b^m	
Distribution	U[15, 20]	U[6, 8]	U[0.1, 0.2]	

For supply and demand, we use the distributions shown in Table 2.2 for the

first period. The remaining periods' supply and demand are randomly generated by altering the first period numbers by ±25% and ±5% for supply and demand, respectively. In addition to these, unit biofuel holding cost (h^f) and unit biofuel transportation cost (r^f) are taken as \$0.05 and \$0.001, respectively. Transportation discount factor (δ) is taken as 0.08. Our test instances can be found online at `http://ise.tamu.edu/LNS/bioenergy_logistics_data.htm`.

2.5.2 Numerical Results

Table 2.3 shows B&C and BD runtimes and optimality gaps for all the data classes. Notice that, when B&C is employed, there is at least one instance in all the data classes except C2 and C5 that cannot be solved until a 2% optimality gap is reached within two hours runtime. Moreover, any of the 80 instances in data classes C9-C16 cannot be solved by B&C and most of the instances in data classes C3, C4, C7 and C8 cannot be solved within two hours by B&C; hence, the reported average run times for these data classes are very high.

Table 2.3 shows the BD results in which the runtimes for the BD-Callback implementation are reported. All the instances in data classes C1-C14 are solved within two hours and significantly faster than the B&C approach. There are total of 4 instances in C15 and C16, that cannot be solved within two hours until 2% optimality gap, however, the average gap for these classes are only 2.2% and 2.9% respectively.

When compared to a base BD implementation, for data classes C1-C8, we observed on average a 14.4% runtime increase when BD-Callback approach is implemented. However, the runtimes for these data classes are already very small with an of average 206 seconds. Therefore, the increase does not have significant effect on the overall solution performance. For larger data classes C9-C16, on the other hand, we observe on average a 13.7% runtime reduction with the BD-Callback implemen-

Table 2.3: Overall results including all BD runs with 2% optimality gap stopping criterion

Class	B&C Results						BD-Callback		
	Runtime (secs)			Opt.Gap (%)			Runtime (secs)		
No.	Ave	Max	Min	Ave	Max	Min	Ave	Max	Min
C1	2630	7200	1135	2.3	5.8	2.0	88	748	45
C2	1909	3932	1150	**2.0**	2.0	2.0	128	166	71
C3	6968	7200	5448	20.6	85.8	2.0	188	275	117
C4	6660	7200	2504	11.9	87.6	2.0	309	477	169
C5	3201	5193	2065	**2.0**	2.0	2.0	155	221	58
C6	3898	7200	2149	9.2	75.4	2.0	217	321	120
C7	6820	7200	4357	17.6	83.5	2.0	278	495	128
C8	6787	7200	4872	26.6	83.6	2.0	289	573	118
C9	7200	7200	7200	38.3	79.9	4.2	583	708	451
C10	7200	7200	7200	20.6	79.2	4.2	957	1384	761
C11	7200	7200	7200	82.8	87.0	77.2	1317	2135	721
C12	7200	7200	7200	83.9	86.7	80.3	2484	3701	1408
C13	7200	7200	7200	14.1	30.4	6.4	1332	1978	810
C14	7200	7200	7200	70.2	72.4	68.3	2674	4765	1418
C15	7200	7200	7200	29.0	34.4	23.2	4219	7202	1904
C16	7200	7200	7200	78.8	81.1	75.1	5300	7202	3037

tation. Moreover, some of the instances in C15 and C16 which cannot be solved by the traditional BD, are solved using BD-Callback approach. Hence, we are convinced that implementing the callback approach for BD algorithm is beneficial especially when solving large data sets.

Moreover, the instances in data classes where the number of candidate biorefinery locations ($|\mathcal{K}|$) is high seem to be more challenging than the other instances. Hence, we can say that number of candidate biorefinery locations plays an important role on the problem's complexity.

2.6 Data Gathering for the Case Study

In this section, we present the source and the nature of data utilized in our case study for the state of Texas. All data is maintained and handled using the GIS (Geographical Information System) software ArcGIS v.10 by ESRI. Data for the case study settings can be found online at `http://ise.tamu.edu/LNS/bioenergy_logistics_data.htm`

2.6.1 Biomass Supply in Texas

We consider five different biomass types in our analysis including switchgrass, forest residues, primary mill residues, urban wood, and crop residues. Forest residues, mill residues, and urban wood are called woody biomass [3]. Corn residues are the agricultural residues from crops like wheat, corn, grain, etc. Switchgrass is a dedicated energy crop which is specifically grown for biofuel production.

For every county in Texas, we obtain the data for all biomass types, except switchgrass, from National Renewable Energy Laboratory 2008 biomass data [38]. In our model, we assume that each county's centroid represents the supply points for the biomass it provides. As the supply amounts, we consider 80% availability for woody biomass (forest residues, mill residues, and urban wood) and 30% availability for crop residues since some crop residues have to be left on the crop land for agricultural reasons. To calculate switchgrass supply for a given county, we follow a specific procedure as follows: We first determine the total crop land from 2007 USDA Census of Agriculture [56]. Assuming that 3.0% of this area is dedicated to switchgrass and 6.25 tons of switchgrass yield per acre, we calculate an estimated average switchgrass yield for each county [21]. Once the base biomass supply amounts are found as described for the first period (base amount) of a planning horizon, we generate estimates of supply amounts for the remaining periods via randomized variations of

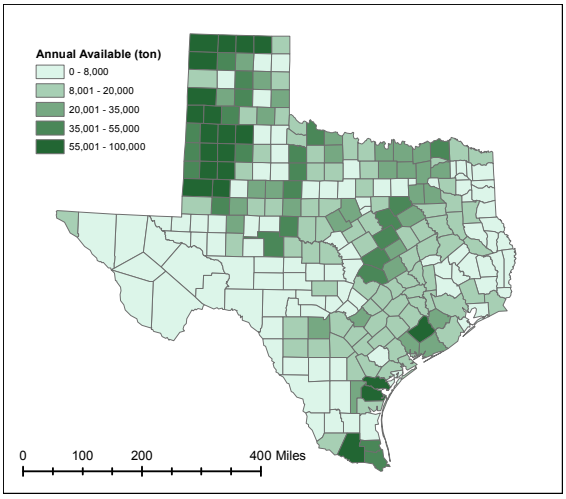
the base supply amounts. This approach allows us to incorporate the time-varying nature of the biomass supply.

It is interesting to observe that, in Texas, biomass supply has regional variation. Figures 2.2a–2.2f show the county-based supply for each individual biomass type as well as the total biomass supply in the state of Texas. Supplies for forest residues and mill residues are located in East Texas where forests are primarily located whereas switchgrass and crop residues supplies are found in Northwest Texas where croplands are primarily located. Urban wood supply, on the other hand, is mostly in population centers.

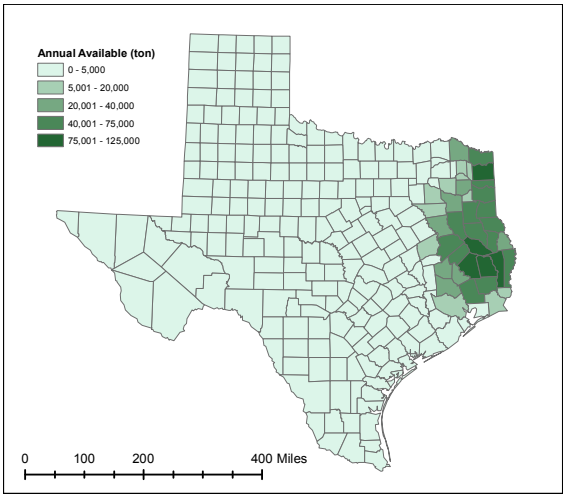
2.6.2 Data on Biomass Types

For our computational studies, we gather biomass type specific data from several resources in the literature and industry/government publications. Whenever a specific data is not available at any of these resources, we use estimates based on available data. Table 2.4 shows the parameters and their values for biomass types. Below, we explain how each parameter value is specifically obtained.

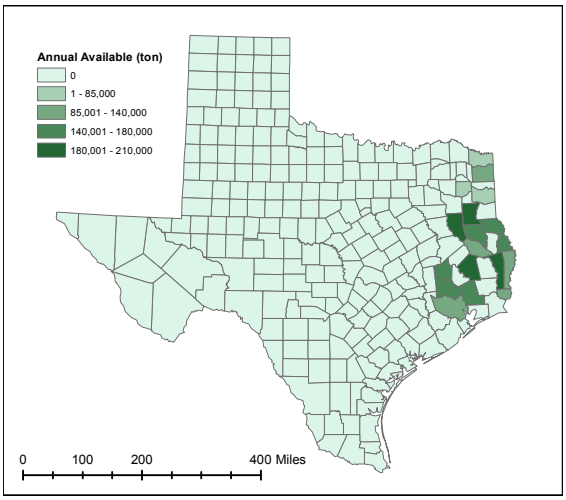
Conversion rates from biomass to ethanol are calculated using "Theoretical Ethanol Yield Calculator" provided by United States Department of Energy [59]. Weight percentages and compositions that are necessary for this calculation are obtained from the "Biomass Feedstock Composition and Property Database" [60]. Due to variations in technologies used for biofuel conversion, Hamelinck et al. [25] state that efficiencies of theoretical conversion rate can vary in the 35-68% range. Taking into account recent technological advances, we assume a base value of 80% as the efficiency in theoretical conversion rates. For example, based on the calculator, theoretically, 96.7 gallons of ethanol can be produced from 1 dt of switchgrass. However, due to adjustment for efficiency, we assume that 77.36 (0.80*96.7) gallons of ethanol is



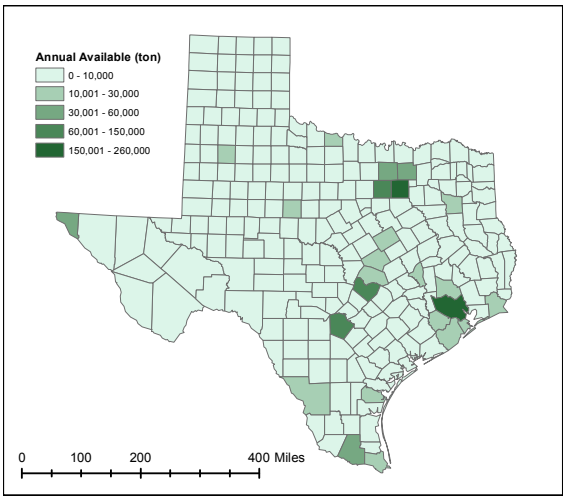
(a) Switchgrass



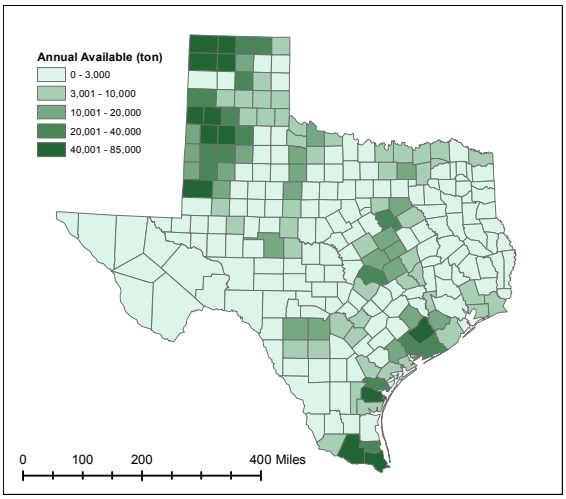
(b) Forest Residues



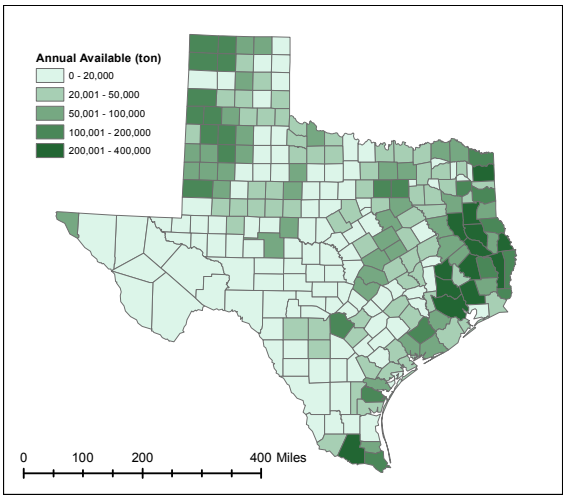
(c) Mill Residues



(d) Urban Wood



(e) Crop Residues



(f) Total Biomass

Figure 2.2: Biomass supply in Texas (scale for supply values in the legends differ for each biomass type)

Table 2.4: Parameters related to biomass types

	Biomass Type				
Parameter (Unit)	Switchgrass	Forest Residues	Mill Residues	Urban Wood	Crop Residues
Theoretical conversion rate (gallon/dt)	96.7	81.5	100.8	98	113
Deterioration factor (%)	5	6	6	6	5
Processing cost ($/dt)	17.406	14.67	18.144	17.64	20.34
Biomass holding cost ($/dt)	5	8.05	8.05	8.05	4.90
Biomass transportation cost ($/dt/mile)	0.13	0.125	0.125	0.125	0.143
Biomass purchasing cost ($/dt)	55	25	30	20	35

obtained from 1 dt of switchgrass in our model.

Deterioration factor denotes how much biomass quality depreciates annually when not processed. For switchgrass, assuming a round baling type and covered facility, the dry matter loss can be estimated as 4.6% in 300 days [46]. Hence, we consider 5% deterioration rate for switchgrass annually. For woody biomass (forest residues, mill residues, and urban wood), we use 6% annual deterioration rate [31]. For crop residues, we assume 5% annual deterioration rate, same as switchgrass.

Processing costs for different biomass types are estimated using the cost of producing a gallon of ethanol and conversion rates. Producing a gallon of cellulosic ethanol costs $2.25 [24]. Moreover, an economic analysis done by APEC (Asia-Pacific Economic Cooperation) shows that 13% of that total cost of producing ethanol is due to facility operations/maintenance. Hence, in this study we estimate 10% of the total cost goes to operations. For example, processing cost for switchgrass is calculated as the 10% of the cost of producing ethanol from 1 dt of switchgrass, i.e., 0.10 * 77.36 gallons/dt * $2.25/gallon= $17.406.

Schnepf [46] claims that the storage cost estimates from the literature is between \$2 and \$17 per ton of biomass. Holding (inventory) costs for corn residue and for woody biomass are provided by Eksioglu et al. [18]. For switchgrass, \$5 per ton is estimated.

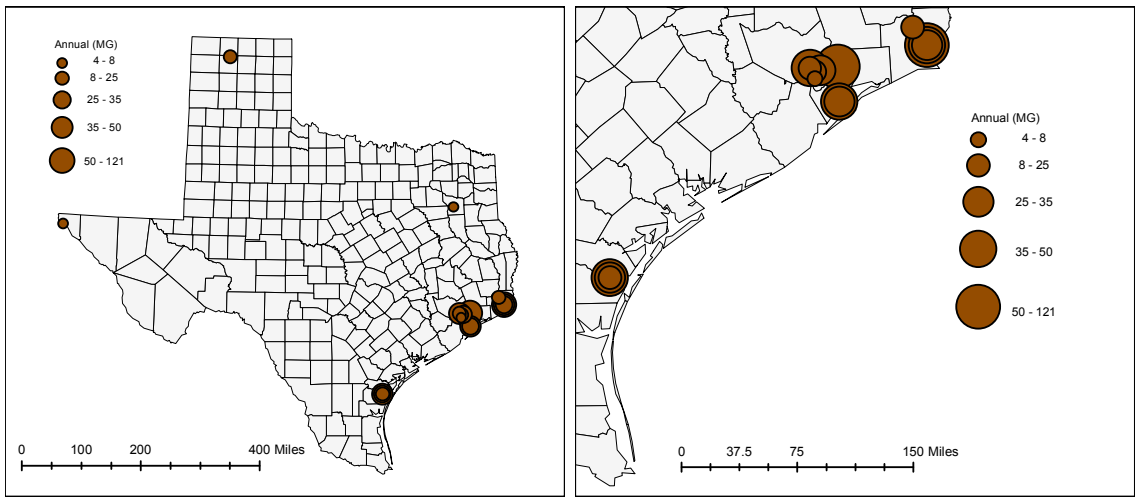
Estimates for transportation cost of different biomass types are taken from the literature. Schnepf [46] observes that the transportation cost estimates in the literature vary between \$0.09 and \$0.63 per ton of biomass. Eksioglu et al. [18] use the transportation cost for woody biomass as \$0.125 per ton, and for corn stover as \$0.143 per ton (for transportation between 25-100 miles). For switchgrass, we estimate a transportation cost of \$0.13 per ton.

Although our model does not explicitly consider biomass purchasing cost, it can be easily added by modifying the cost coefficients of the first and third terms corresponding to out-of biomass source transportation costs in the objective function (2.1). Specifically, letting p_b denote the purchasing cost for biomass type b, in the term with variable X^1_{ijbt}, the cost coefficient is modified as $(r^m_b d_{ij} + p_b)$ and, in the term with X^3_{ikbt}, the coefficient is modified as $(r^m_b d_{ik} + p_b)$. This simple modification neither changes the MIP model nor the BD based solution algorithm. For our case study analysis, purchasing costs for each biomass type are taken from United States Environmental Protection Agency [66] and Schnepf [46]. The numbers are approximated according to the given ranges in these sources.

2.6.3 Ethanol Demand in Texas

Demand for biofuel is essentially at the petroleum refineries where the blending of biofuel and gasoline takes place. Thus, to calculate the biofuel demand, we first identify the existing petroleum refineries in Texas. As of 2012, there are 26 operational oil refineries in Texas and 18 of those produce gasoline. Gasoline production capacities

and the locations of these refineries are obtained from the Annual Refinery Report of EIA [63]. Hence, we consider these 18 facilities as our demand points. We assume 50% of the total capacity in these refineries utilized to satisfy demand in Texas with E-10 (10% Ethanol) gasoline-ethanol production. We obtain the ethanol demand in each of the blending facilities using these input values. The results indicate that, the total ethanol demand in Texas is estimated to be around 710 million gallons annually. We note that, with this calculation, we obtain the base demand amount at the biorefineries. Since we consider a multi-period planning horizon, similar to determining the supply values, we obtain demand for the remaining time periods via perturbing the base demand values.



(a) Demand locations and their amounts in Texas

(b) Demand locations and their amounts in the Gulf area

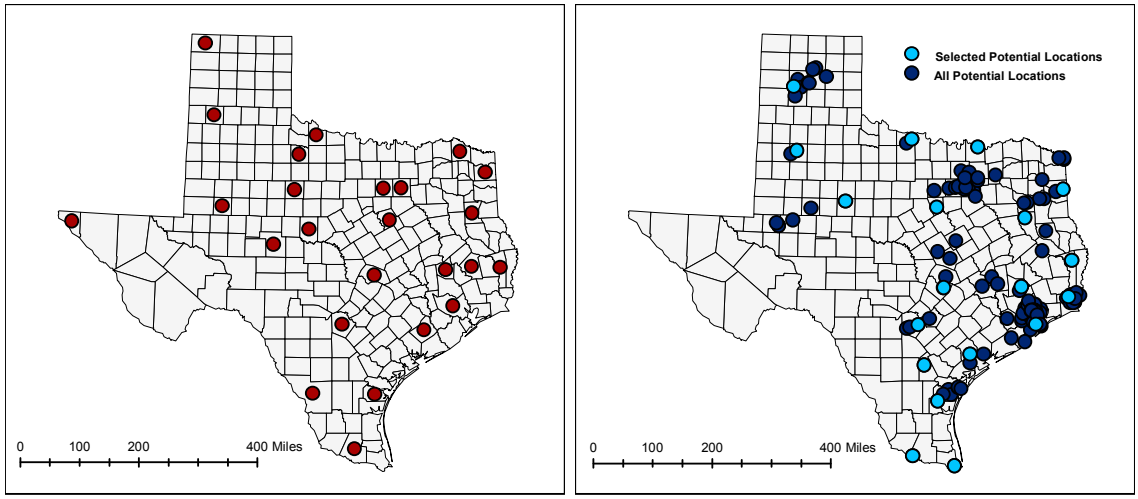
Figure 2.3: Biofuel demand in Texas

Figure 2.3a shows the demand locations and their amounts in Texas. As shown, most of the demand points are located at the Gulf area. Figure 2.3b provides a detailed view of the demand points in the Gulf area. As for the biofuel related parame-

ters, we assume the holding and transportation costs for biofuel as 0.05 $/gallon/year and 0.001 $/gallon/mile, respectively, as suggested by Eksioglu et al. [18].

2.6.4 Candidate Biorefinery and Collection Facility Locations

In order to determine the candidate locations for collection facilities, we use Texas Department of Transportation (TxDOT) regional system. TxDOT divides the state of Texas into 4 regions and 25 districts [53]. We select one county with the highest total biomass from each district. The center points of those selected counties are chosen as the potential collection facility locations in our study. We follow this approach to achieve dispersed candidate locations. Figure 2.4a shows the selected candidate locations for collection facilities.



(a) For collection facilities (b) For biorefineries

Figure 2.4: Candidate facility locations

To determine candidate biorefinery locations, we use data presented by EPA [67]. Of all the candidate locations provided therein, we consider only the ones that are suitable for biorefinery construction with an "excellent" or "outstanding" potential

rating. Among those, 19 candidate biorefinery locations, that are at least 60 miles apart from each other, are selected to generate a disperse candidate location set. Figure 2.4b shows all the "excellent" and "outstanding" candidate biorefinery locations as well as the selected candidate locations. Although currently there are four existing biorefineries in Texas with capacity of 375 MGY [19], in this study, we do not consider these biorefineries. Nevertheless, our mathematical model is capable and flexible enough to incorporate these biorefineries.

2.6.5 Fixed Investment Costs

USDA (U.S. Department of Agriculture) estimates the total investment cost of a 40MGY capacity biorefinery as $320 million [57]. Moreover, Wallace et al. [72] estimate that doubling the size of a biorefinery increases the investment cost by 60%. The fixed costs associated with biorefineries in this case study are calculated using these figures. For the collection facilities, we use the fixed cost estimates in [1] which provide fixed cost estimates for an indoor anaerobic storage facility.

We consider 10 years of project life and 10% interest rate in this study. The IRS' MACRS depreciation rate numbers are used to estimate the salvage value of the biorefineries at the end of the project life [68]. These numbers are then used to calculate the present net fix cost of biorefineries. Lastly, we consider four alternative capacity levels, denoted as *L1*, *L2*, *L3*, and *L4*, for a biorefinery as well as for a collection facility. Table 2.5 summarizes the cost and capacity data for varying size biorefineries and collection facilities.

2.7 Network Analysis Based on Varying Input Parameters

In this section, we analyze a wide spectrum of realistic cases based on industrial data and geographical/logistical data from the state of Texas described in the previous section. In this computational analysis, we consider five different biomass types

Table 2.5: Parameters related to facilities

	Size Level				
Parameter	*L1*	*L2*	*L3*	*L4*	Unit
Production capacity	40	80	120	160	million gallons
Biorefinery fixed cost	320	512	665.6	819.2	million $
Collection facility fixed cost	2	3.5	4.5	6	million $
Collection facility inventory capacity	200	500	700	1,000	dt (thousands)
Biorefinery biomass inventory capacity	10	25	35	50	dt (thousands)
Biorefinery biofuel inventory capacity	10	20	30	40	million gallons

with varying associated transportation/processing costs, deterioration and conversion rates and examine the network and solution effects of variations in conversion rates (based on technology), transportation discount factors (under transportation economies of scale), supply variability, demand variability, and transportation and production costs.

Building on the base setting S1, for analysis, we generate nine additional different settings by varying input parameters values. In the first six settings, S2-S7, we vary one input parameter in S1 at a time to generate the settings as shown in Table 2.6.

Table 2.6: Settings for analysis

Parameters	S1	S2	S3	S4	S5	S6	S7
Efficiency in conversion rate (%)	80	100	80	80	80	80	80
Transportation discount (%)	8	8	0	8	8	8	8
Supply variability (±%)	5	5	5	50	5	5	5
Demand variability (±%)	1	1	1	1	20	1	1
Transportation costs increase (%)	0	0	0	0	0	20	0
Production costs decrease (%)	0	0	0	0	0	0	20

In setting S2, we only change the conversion rates for all the biomass types. Instead of an 80% conversion rate efficiency, we consider 100% efficiency to better reflect technological developments in processing biomass and its impact on logistics networks. In setting S3, we eliminate transportation discount factor between collection facilities and biorefineries to examine the effects of transportation economies of scale on network structure. In settings S4 and S5, supply and demand values change more drastically from period-to-period in the planning horizon, respectively. Since biomass supply and demand can be variable, these settings help us to examine the effects of variability on network design. For setting S6, we examine the increase in the unit transportation costs, via increments by 20% for all biomass types, as some sources in the literature suggest higher per-unit per-mile transportation costs. In setting S7, the unit production costs are decreased by 20% for all biomass types due to expected positive impacts of continuing technological advances on production efficiencies. In setting S8, we consider only woody biomass types (forest residues, mill residues, and urban wood) as available biomass supply and, in setting S9, we consider only switchgrass and crop residues as available biomass supplies. Lastly, in setting S10, we consider the seasonality in the supply and analyze its effects on the network structure.

2.7.1 Analysis Results

The BD-Callback algorithm presented in Section 2.4 is used to solve all the settings mentioned above until 0.5% optimality gap is reached. We present the results and analysis under three groups for settings S1-S7, S8-S9, and S10.

Settings S1 to S7 - Changing Input Parameters

Figure 2.5 shows the facility locations for the base setting **S1**. In this case, there are a total of five open biorefineries: three large ($L4$), and two mid-large

(*L3*). All three *L4* size biorefineries are located in East Texas where majority of the large demand points and almost all the forest residue and mill residue biomass are located. The *L3* size biorefineries are open in the northwest and south parts of the state due to the existence of demand in those regions as well as switchgrass and crop residue supplies. Moreover, there are two large *L4* size collection facilities open at the boundary of north-west and the south-east parts of the state (mainly Central Texas) to facilitate transfer of biomass mostly towards southeast Texas where the major demand exists. In addition, there are two *L4* size collection facilities open in North Texas.

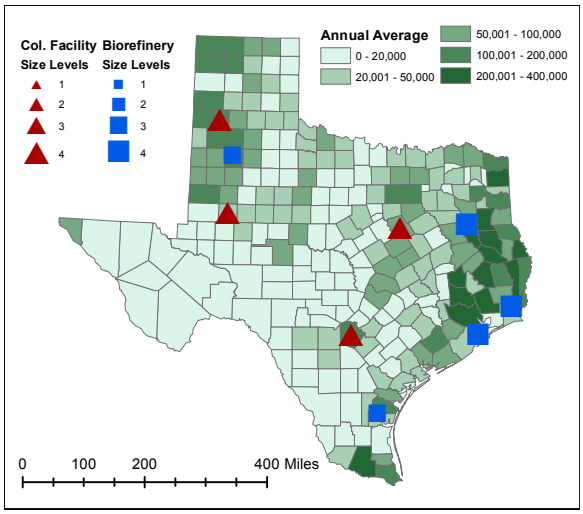


Figure 2.5: Biorefinery and collection facility locations for S1 - Base setting

Figure 2.6a shows the biorefinery and collection facility decisions for **S2**. Biorefinery location decisions are slightly changed from S1. Since conversion rates are higher in this setting, more biofuel is produced to satisfy most of the demand with available biomass supply in East Texas. Additional biomass supply is obtained mostly from urban wood in Central Texas and crop residues in the northwest which are collected

at the new facilities opened in Central Texas. Thus, while the locations of four of the biorefineries remain the same as in S1, an additional *L2* size biorefinery is opened in East Texas to support increased production and the size of the biorefinery in northwest is reduced to *L1* to serve the demand in the north and west with biofuel produced from switchgrass locally collected.

Facilities locations for **S3** are shown in Figure 2.6b. We observe that while the same biorefineries are opened as in S1, no collection facilities are opened. This is simply because, in this setting, no biomass transportation cost saving is available between collection facilities and biorefineries. Hence, all of the biomass shipments are direct shipments from supply points to biorefineries.

The facilities open in **S4** are given in Figure 2.6c. Notice that average supply amounts for the counties are different in S4 (i.e., for some counties the shading is darker or lighter in the map) since the supply variability is higher. In the solution for S4, only one of the biorefinery locations is different than the base setting S1. Three of the opened biorefineries, all in *L4* size, are located in East Texas. The remaining two biorefineries are located in Northwest and South Texas as in S1. To adapt to the supply variability, five collection facilities, that hold inventory, are opened. These collection facilities are opened mainly in Central Texas, between East and Northwest Texas where the demand and most of the supply are located.

Figure 2.6d shows the open facilities for **S5** when demand variability is high. Biorefinery location and size decisions are similar to the S1 and S4 settings. Since demand is higher in the southeast and the biorefineries can hold biomass as well as biofuel inventory, to accommodate the demand variability, we observe an increase in the size of the biorefinery opened in South Texas. Five collection facilities are opened, however, in this setting they are generally more closer to the biorefineries which respond to changes in demand that is mainly in the Southeast Texas. The

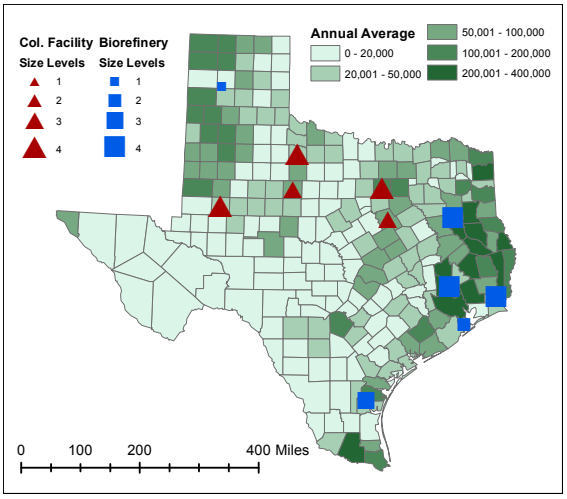
total demand for this setting turned out to be lower than the total demand in S1 due to random data generation for demand variability.

In **S6** in which the unit transportation costs are higher, five biorefineries, three *L4* size and two *L3* size are opened similarly to other settings as shown in Figure 2.6e. In addition, eight collection facilities of which six *L4* size and two *L3* size are located in Central as well as Northwest and East Texas. This helps with the transportation of crop residue and switchgrass to the biorefineries in East Texas. With the increase in unit biomass transportation cost, the role of collection facilities becomes more important since they provide higher transportation discounts. Thus, both the number and the spread of collection centers in the state increase significantly so that they are closer to supply locations.

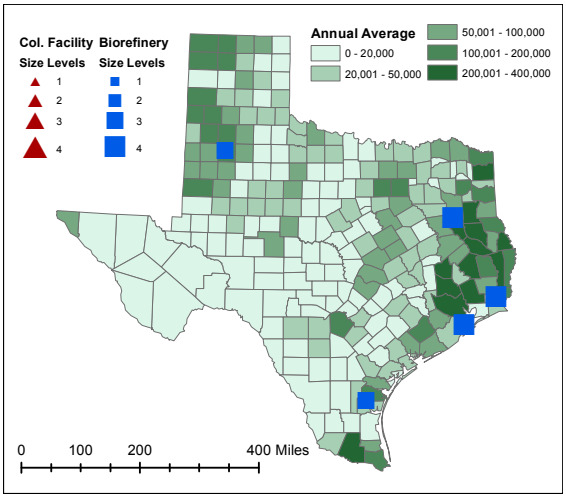
Figure 2.6f shows the facilities decisions when unit production costs are lower in **S7**. Biorefinery locations are very similar to the base S1 setting, only one biorefinery location in East Texas changes slightly. However, six collection facilities, with five of them being *L4* size, are now opened around Central Texas. Lower production costs motivates more processing which is supported by transferring higher amounts of biomass to the biorefineries. The higher transfer amounts at a lower transportation cost are achieved due to transportation economies-of-scale by opening more spread-out collection facilities.

Settings S8 and S9 - Biomass Groups

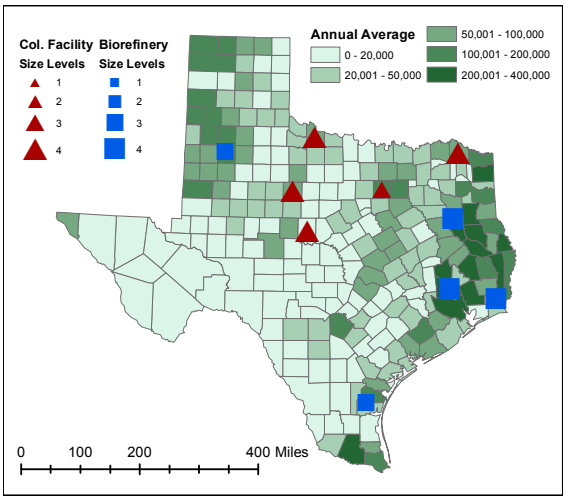
In **S8**, only woody biomass types for which the majority of the supply located in East Texas are considered as shown in Figure 2.7a. Two *L4* size biorefineries are opened in that region to satisfy the demand. Another *L1* size biorefinery is opened in South Texas. Moreover, three collection facilities are opened in East and Central Texas to facilitate transportation of urban wood and forest residue. Notice that



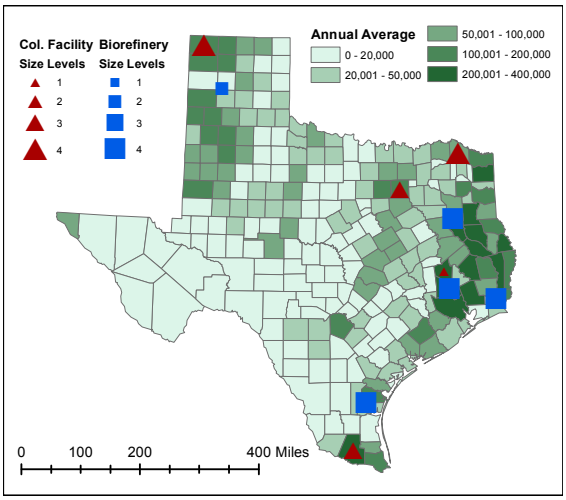
(a) S2 - Increased conversion efficiency



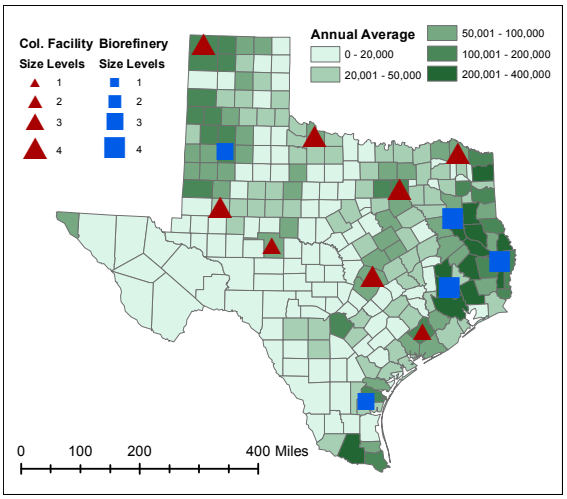
(b) S3 - No transportation cost discount



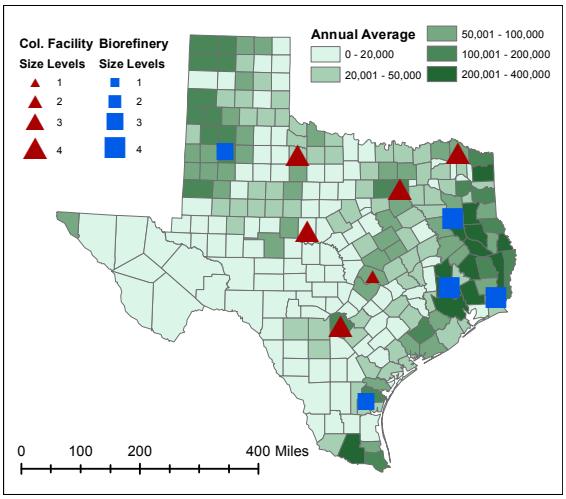
(c) S4 - High supply variability



(d) S5 - High demand variability



(e) S6 - Higher transportation costs

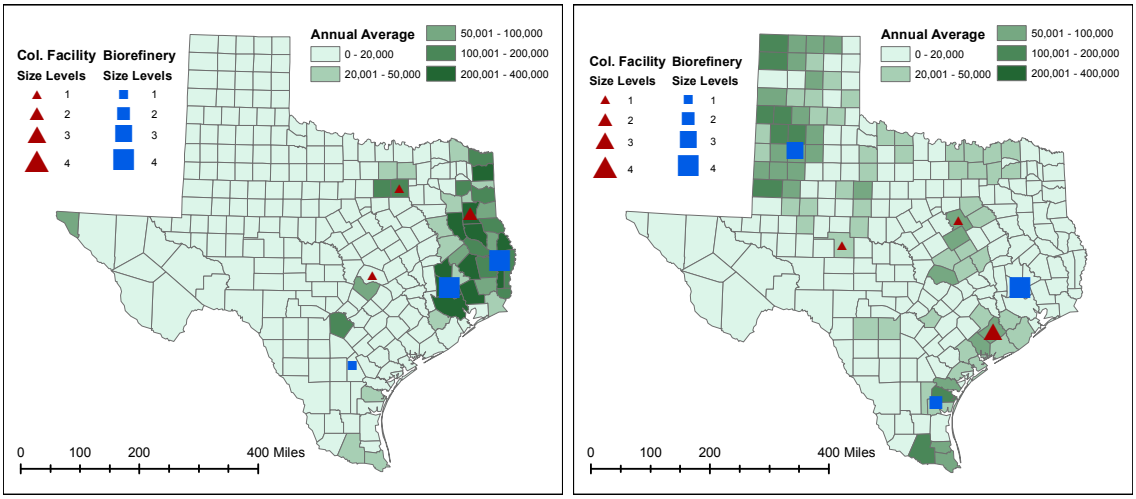


(f) S7 - Lower production costs

Figure 2.6: Biorefinery and collection facility locations for S2-S7

there are no biorefineries open in North or West Texas. This means that the demand points in these locations are satisfied directly from the biorefineries opened in East and South Texas. Since supply and demand in this special case is concentrated in one area, East Texas, biorefineries are also located in the same area for direct transportation of biomass to biorefineries and biofuel to demand points.

In **S9**, we consider only crop related biomass types as supply sources, and thus, as shown in Figure 2.7b, biomass sources are primarily located in North Texas and some in the South Texas. One *L4*, one *L3* and one *L2* size biorefineries are opened in East, North and South Texas, respectively. Moreover, three collection facilities are established in East and Central Texas. Supply is more dispersed in this setting when compared to S8; hence, collection facilities are located more spread out. Biorefineries are also dispersed as it is generally cheaper to ship biofuel instead of biomass.



(a) S8 - Only Forest Residue, Mill Residue and Urban Wood supply

(b) S9 - Only Switchgrass and Crop Residue supply

Figure 2.7: Biorefinery and collection facility locations for S8-S9

Setting S10 - Seasonal Supply

Biomass supply, especially from crop related biomass, typically show seasonality. Therefore, in **S10**, we consider seasonal supply for this type of biomass. We employ the same data in base setting S1, but, in this case, each time period corresponds to a season rather than a year. We assume that switchgrass and crop residues are harvested only in Summer and Fall as is usually the case and that woody biomass supply is available in all seasons as considered in S1.

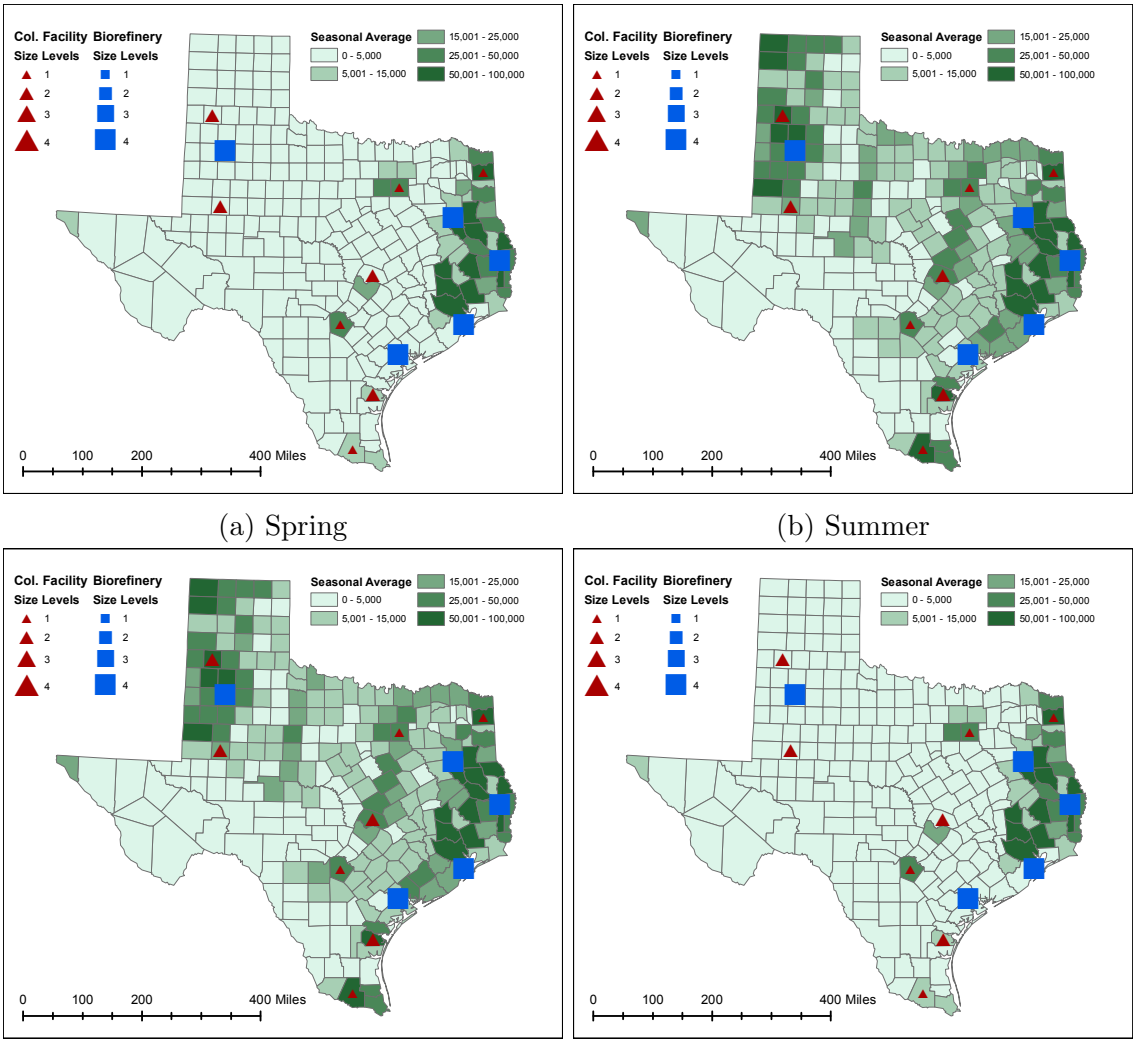
Figure 2.8 shows the facility locations for S10 as well as seasonal biomass supply. We observe that, in Spring and Winter, the biomass supply is concentrated only in East Texas where woody biomass is found. During Summer and Fall, biomass supply is widely available from all sources as in S1. The locations of biorefineries and collection facilities are still shown for all seasons in Figure 2.8. A total of five biorefineries along with eight collection facilities are opened. All five biorefineries are *L4* size and they are found in the North, South, and the East to serve demand in the North-and-West, South-and-East, and the East, respectively. The number of opened collection facilities are higher than the S1 setting while their sizes are smaller. This is primarily because more biomass inventory is needed in different regions to satisfy the demand in Spring and Winter.

2.7.2 Discussion on Results

In all of the settings in our network analysis, we observe a consistent pattern in terms of biorefinery locations. The large biorefineries are located in East Texas where most of the demand as well as most of the biomass supply (as woody biomass) exist. Hence, it seems that East Texas is the most attractive area to locate biorefineries. Smaller biorefineries are also located following a pattern of one in the South again to serve to some of the demand in the East and to supply the gasoline refineries in their

vicinity, and another in the Northwest to process crop residue and switchgrass and supply the refineries in the North and the West Texas. This means that the solution structure in terms of biorefinery locations is relatively robust without significant changes even when some of the system input parameters are significantly altered.

Locations and the size of collection facilities, on the other hand, vary significantly among the settings. However, as a general pattern, we notice that in majority of the settings, collection facilities are located in and around Central Texas which is the area between woody biomass rich region in the East Texas and crop-related biomass region in the North Texas. These collection facilities connect these two important areas and provide opportunities for transportation discounts due to economies-of-scale. Indeed, in the settings in which supply and demand variabilities are not high, the collection facilities are still required even if they do not hold significant inventories to respond to period-to-period variations. In these settings, they act as consolidation locations for the pass-through biomass on its way to biorefineries. On the other hand, in settings S4, S5, and S10, collection facilities are also utilized to store inventory so that the variations in supply or demand is handled in a cost effective manner.



(a) Spring (b) Summer

(c) Fall (d) Winter

Figure 2.8: Biorefinery and collection facility locations for S10 - Seasonality in supply amounts (Same set of locations are shown for each season)

The major reason why biorefinery locations are robust to parameter changes is the fact that a very large portion of the biofuel demand is concentrated in East Texas. Moreover, woody biomass (specifically forest residues and mill residues) are located in the same area. Thus, being close to both biofuel demand locations and a significant percentage of the biomass supply with good conversion efficiency and

lower purchasing cost deem East Texas a very attractive place for biorefineries. As a result, locations of these facilities do not change significantly although problem parameters are altered. However, collection facility locations lack this robustness because they serve a different purpose including both inventory holding and facilitating transportation economies-of-scale. Specifically, when there is low variability in biomass supply and biofuel demand, the main purpose of collection facilities is to provide transportation discount. Therefore, if problem parameters affecting the biomass transportation (either quantity or unit cost) are altered, collection facility locations change accordingly.

We also observe the following regarding system costs. Among the settings S2-S7, the fixed costs share of the total system cost is lowest in S6 with only 35% and highest in S2 with 44% since less biomass is needed due to increased conversion efficiency leading to a decreased total transportation cost. Total system cost reaches its highest in S6 due to unit transportation cost increase and it reaches its lowest in S2 when conversion rates are increased facilitating more efficient demand satisfaction with less biomass. For settings S8 and S9, the total system costs are smaller when compared to other settings since the settings consider smaller scale instances. In S10, in which we examine seasonality in supply explicitly, all of the cost components increase when compared to S1, resulting in a 15% increase in the total system cost. In particular, total transportation and production costs are increased by 23% and 10%, respectively. Moreover, total inventory cost is also increased, especially for the biomass inventory in collection facilities. As a result, in S10, fixed costs share in the total system cost decreases to 25% which is the lowest among all settings.

Finally, settings S8 and S9 can be treated as the decomposition of the base setting S1 in terms of geographical area and input data. Thus, it is interesting to observe that if the individual solutions to S8 and S9 are put together to prescribe a solution for

S1, the resulting network structures are quite different in locations and capacities of both biorefinery and collection facilities. Furthermore, the total objective function value of S8 and S9 solutions is 8% more than the objective value of S1 solution. Thus, an integrated solution that considers the overall region of interest provides a significantly more effective bio-energy logistics network structure.

2.8 Concluding Remarks

In this section, we present a deterministic multi-period and multi-biomass network design model that covers both strategic and tactical level decisions in the upstream and downstream echelons of bio-energy supply chain. Our model simultaneously incorporates design and planning decisions and application specific system constraints such as time-dependent biomass deterioration, transportation economies-of-scale, locations of both collection facilities and biorefineries as well as their capacities and inventory decisions. Although there are other studies presenting similar models for bio-energy supply chains, our study provides a comprehensive mathematical model which effectively bring together specific application features while addressing key cost components in bio-energy supply chains. The approach then provides the ability to capture the dynamics and trade-offs among important design and operation characteristics in a cost effective manner, and therefore, it is useful in analysis for decision making in the developing area of bio-energy supply chains.

As we highlight in Section 2.3 and also as noted in the current literature, bio-energy supply chain literature lacks efficient solution methodologies to allow the optimal solutions for realistic size problems for comprehensive analysis in this developing area. To satisfy this need in the literature, we present a BD based solution algorithm that can solve large-scale instances effectively and efficiently. Our computational results on a wide spectrum of test instances show that our solution algorithm provides

significant computation time reduction when compared to branch-and-cut approach (with a state-of-the-art implementation in off-the-shelf software). In this context, to improve the efficiency of the BD approach via strong lower bounds, we also develop surrogate (valid) constraints by taking into account the interrelation among the decisions on multi-period inventory levels, facility capacity levels, biomass to bio-fuel conversion factors. This solution algorithm can also be applied to similar multi-tier, multi-period, and multi-product (without product conversion) supply chain network design problems to simultaneously determine locations and capacities of facilities in multiple levels as well as transportation, production and inventory decisions that are modeled as large scale MIPs.

Moreover, we demonstrate the effectiveness and applicability of our model through a case study conducted using realistic data from the state of Texas managed in a GIS (Geographical Information Systems) environment. We perform analysis on the case study based on changing some of the input parameters. Specifically, we consider five different biomass types with varying associated transportation/processing costs, deterioration and conversion rates and significant geographical diversity on their availability. We examine the network and solution effects of variations in conversion rates (based on technology), transportation discount factors (under transportation economies of scale), supply and demand variability, and transportation and production costs. We observe that while the biorefinery locations are relatively robust to changes in input parameters, collection facilities are quite sensitive. Furthermore, seasonality in supply is handled mainly via increased inventory levels and overall increase in inventory, production, and transportation costs. Finally, we also observe that decomposing the geographical region of study based on the locational clustering of biomass supply and addressing network design for each cluster separately produces suboptimal results. Integration in terms of biomass types and geographical area for

a comprehensive analysis, although increases the size and complexity of the problem, provides improved cost effectiveness in the order of 8% in our case study. This case study is the first generalized and state-wide biofuel/biomass supply chain study done in Texas. As noted before, investment in bioenergy in Texas is attractive and sensible due to its proximity to oil industry (demand points) and energy facilities, agricultural land availability (in and around the state) and a vibrant economy.

3. STOCHASTIC BIOMASS LOGISTICS NETWORK DESIGN UNDER PRICE-BASED SUPPLY

3.1 Introduction

Bio-energy is renewable energy made from organic matter. In fact, people have been using bio-energy ever since they started burning wood to provide heat to stay warm and to cook. Nowadays, bio-energy provides more than just heat but electricity, fuel and other types of energies that we need in our daily lives. Biofuels, such as ethanol and biodiesel, are some of the most common bio-energies used these days.

Renewable Fuel Standard (RFS) program regulations require that transportation fuel sold in the United States contains a minimum volume of renewable fuel, i.e., biofuel. This program was extended in recent years to increase the volume of renewable fuel required to be blended into transportation fuel from 9 billion gallons in 2008 to 36 billion gallons by 2022 [65]. Moreover, ethanol blend wall, currently 10%, is very likely to increase to 15% or more in near future. This will create more demand for ethanol and hence, an expansion in biofuel industry. These clearly show that the investment to biofuels increases substantially in the coming years.

Previous research indicate that biomass supply cost accounts for about 20-40% of the total biofuel production cost and about 90% of the costs of supplying biomass is related with biomass logistics [18]. Therefore, in this paper, we focus on the biomass logistics and the upstream of the biofuel supply chains while considering energy crops, e.g., switchgrass, as biomass source.

More energy crops such as switchgrass, must be produced to satisfy the biofuel industry's growing supply needs. However, studies like Jensen et al. [28] and Villamil et al. [71] indicate that most farmers do not have enough knowledge about energy

crops to adapt them. Moreover, the uncertainty about the financial viability of energy crops is a limiting factor for farmers to adapt energy crops. Therefore, incentives should be given to farmers to encourage them to produce these crops. One way to achieve this is to offer economic benefits to farmers. Okwo and Thomas [39] consider wholesale and acreage contracts to give incentives to farmers to adapt switchgrass production. A wholesale contract guarantees the farmers to sell all their products based on the determined contract price value.

In this study, we use a policy where farmers are offered a unit biomass price by the biofuel producer. Our objective is to determine the best price to offer to a set of farmers by using their land allocation decision model as input. In addition to this, we also aim to determine the biomass supply chain network structure to facilitate switchgrass transportation from farms to biorefineries for biofuel production. Therefore, our model includes both pricing and supply chain network design decisions. Moreover, we also incorporate uncertainty in switchgrass yield that is an important concept in biomass-biofuel supply chain design due to weather and other environmental uncertainties in reality. To solve this problem, we propose an algorithm based on the L-shaped method that generates multiple strengthened-cuts at each iteration and we also propose a Sample Average Approximation (SAA) approach to solve large-scale instances within reasonable time. Lastly, we conduct an extensive case study in Texas considering the effects of different problem parameters on the expected total system cost, biomass price and supply chain network structure.

This section is organized as follows: Section 3.2 reviews the recent literature and highlights related papers published in this area. The problem definition and mathematical formulation are presented in Section 3.3. We introduce farmer's decision model and the reformulated stochastic program in Sections 3.4 and 3.5, respectively. The L-shaped method based solution algorithm is presented in Section 3.6. In Sec-

tion 3.7, we discuss the case study problem parameters and how the input data is obtained. Using these data values, we present a computational study in Section 3.8, which includes a comparison of two different cut generation approaches, an SAA analysis and the value of stochastic programming. Our extensive case study, which is conducted in Texas, is presented in Section 3.9 along with a discussion of the outcomes. In Section 3.10, we summarize our results and contributions.

3.2 Literature Review

Several studies on biomass-biofuel supply chain design are conducted and published in recent years. We first start with two review papers related with biomass-biofuel supply chains that are published very recently.

Yue et al. [73] identify the key research challenges and opportunities in modeling and optimization of bio-energy supply chains. The authors categorize the related literature according to the biomass type used such as food crops, cellulosic biomass, algae, and other components including conversion technology, sustainability, uncertainty, multi-objective optimization etc. They claim that multi-scale modeling and optimization approach, which considers four system layers i.e., ecosystem, supply chain, process and molecule, can play an important role in addressing the research challenges in bio-energy supply chains. De Meyer et al. [14] review the growing literature on the bio-energy supply chain design and management focusing on the upstream (biomass logistics). The authors review 71 publications and the selected literature is categorized according to (i) the mathematical optimization methodology used, (ii) the decision level and decision variables addressed, and (iii) the objective to be optimized. The authors state that most of the publications apply mathematical programming approaches, specifically mixed-integer programs (MIPs). They argue that the major drawbacks of these models are their large sizes and long runtimes.

The authors say that the literature lacks of mathematical models that consider economical, environmental and social objectives in an integrated fashion.

Some studies in the literature consider uncertainty in problem parameters such as biomass supply, biofuel demand and biomass and biofuel prices. Different approaches including scenario-based optimization, stochastic programming and robust optimization, are utilized to incorporate these uncertainties. We observe two major issues in the literature specifically in the studies that propose stochastic programs. The first issue is the number of scenarios considered in analysis purposes. Most studies in the literature consider very few number of scenarios and this might not correctly capture the underlying nature of uncertainty, which hinders the very reason to incorporate uncertainty at first place. The second issue is the necessity of efficient solution methodologies. Almost all studies in the literature directly employ off-the-shelf softwares in order to solve the problems they consider, which results in huge computational burden and long runtimes. Only very few studies propose solution methodologies and algorithms that are capable of solving large-scale problems. In addition to these two issues, the importance of biomass price and its effect on the supply chain network design is not sufficiently addressed in the literature. According to our knowledge, there is only one study, Bai et al. [4] that incorporates the effects of biomass price on the biomass-biofuel supply chain network. Below, we summarize Bai et al. [4], which has a deterministic problem setting, and list some of the recent literature on biomass-biofuel supply chain design under uncertainty.

Bai et al. [4] introduce a game-theoretic optimization model that designs the biomass-biofuel supply chain and determines farmers' and biofuel producer's decisions. The authors propose two models: (i) a non-cooperative Stackelberg leader-follower game model (decentralized) and (ii) a cooperative game model (centralized). In their first model, each actor tries to maximize their individual profit. The farmers

have the option to send their biomass to local markets or to nearby biorefineries where they need to pay the shipment costs on their own. Therefore, the farmers' problem is to determine how much flow to send to local markets or opened biorefineries to maximize their profit. On the other hand, biofuel producer determines where to locate its biorefineries and their associated biomass prices to maximize its own profit. This Stackelberg game model is transformed into a mixed-integer quadratic program (MIQP) and solved using off-the-shelf software. The second model assumes a cooperation between all the actors and it is constructed as a mixed-integer nonlinear program (MINLP), which maximizes the system profit. The authors implement both models on a case study in Illinois using networks with 10-20 farms, 10-20 candidate biorefinery locations and 10-20 local markets. They observe that the system profit increases when there is cooperation between the biofuel producer and farmers. When cooperation exists, more biorefineries are open and more land is allocated to produce biomass.

Kim et al. [29] present a two-stage stochastic programming model that maximizes the system profit and determines the location and capacity levels of two level conversion facilities along with the flow decisions between each node in the network. The authors consider various uncertain parameters including availability and acquisition costs of each biomass type, costs of transporting intermediate products, biomass and final products and price of each final product. They determine the five most important uncertain parameters affecting the objective function that are: (i) the price of the final product, (ii)-(iii) the conversion yield ratios of the two conversion processes, (iv) maximum demand and (v) biomass availability. A total of 33 scenarios are generated changing these five scenarios with their high and low values and the expected value scenario. In addition, they use a data set from the Southeast U.S. with 30 supply, 29 (level 1) and 10 (level 2) candidate conversion facilities and 10 demand

points. The authors implement the robustness analysis and Monte Carlo sensitivity analysis to compare the multiple scenario design to the single scenario design.

Chen and Fan [10] propose a two-stage stochastic program to design biomass-biofuel supply chain. The objective is to determine refinery and terminal locations along with biomass and biofuel transportation decisions and biofuel production decisions while minimizing the total system cost. The authors consider two sources of uncertainties that are biomass supply and biofuel demand. Moreover, they propose a progressive hedging (PH) algorithm, in which they decompose the the whole problem with respect to the scenarios. The authors conduct a case study in California with a number of supply points ranging between 14 to 57, 28 candidate biorefinery locations and 143 demand locations. Two settings are considered where demand and supply uncertainties are considered with 4 and 10 scenarios, respectively.

Dal-Mas et al. [12] present a stochastic mixed-integer program to determine where and how much biomass to produce, biomass transportation decisions and biorefinery locations and capacities. The authors propose two alternative objective functions: (i) maximizing the system profit and (ii) minimizing the risk associated with the investment. The biomass purchase costs and biofuel market price are assumed to be uncertain and these uncertainties are addressed by a set of possible scenarios. A case study in Northern Italy considering corn-to-ethanol supply chain is conducted to demonstrate the capabilities of the proposed framework. The case study area is divided into 60 square regions each with 50km of length. Moreover, different modes of transportation such as truck, rail and barge are considered to transport both corn and ethanol. The results suggest that under profit maximization objective, there is always a reasonable probability to obtain profitable results even when the biofuel prices decrease over the years. The results under risk minimization objective show that when the worst case market scenarios are considered, the system becomes

profitable only if the biofuel prices are very high.

Osmani and Zhang [40] and Osmani and Zhang [41] present similar models to design and optimize lignocellulosic based ethanol supply chains. Both papers introduce two-stage stochastic programs considering uncertainties in (i) supply (biomass yield), (ii) ethanol demand, (iii) biomass price and (iv) ethanol price. Although both models maximize expected profit of the supply chain by optimizing the strategic and tactical decisions, Osmani and Zhang [41] simultaneously minimizes carbon emissions as well.

Objective function in Osmani and Zhang [40] includes the revenue from switchgrass and ethanol sales and tax credit from sale of subsidized biofuel, as well as cost components such as marginal land rental, switchgrass cultivation and harvesting, biorefinery annualized capital and operational and switchgrass transportation. The two-stage SP model determines biorefinery locations at the first-stage and all the other tactical decisions, i.e., purchasing and transportation, related to the supply chain at the second-stage. The authors conduct a case study in North Dakota (ND), where they assume all 53 counties of ND are potential biorefinery locations and demand points. Three independent random variables, (i) ethanol price, (ii) annual rainfall level and (iii) ethanol demand are utilized to capture the four uncertainties mentioned earlier. These three random variables are discretized into 10 levels and a total of 1000, i.e., 10^3, scenarios are generated to accurately capture the underlying uncertainty. The authors solve the Deterministic Equivalent Problem (DEP) using off-the-shelf softwares without proposing any solution methodology. They compare the DEP solution with the solution of the deterministic model where all the random variables are assumed to take their expected values. The authors state that although both results give the same optimal biorefinery locations, the stochastic model (DEP) outperforms the deterministic model since the deterministic model underestimates

switchgrass land allocation and ethanol production volume.

Osmani and Zhang [41] propose a similar model considering the same four uncertainty parameters with the addition of carbon emission reduction term in the objective function. In addition to the previous model, authors include another first-stage decision that is the selection of ethanol conversion technology. Moreover, unlike the previous study, the authors utilize Sample Average Approximation (SAA) approach to select and generate the scenarios they use. Furthermore, a case study is conducted in four American Midwest states: Illinois, Iowa, Minnesota and Wisconsin. The authors consider two geographical scales for this case study: (i) cooperation, where four states are considered as a combined region and biomass and ethanol exchange is allowed between the states, and (ii) stand-alone, where each state is considered as an individual region. The results show that each state gives better financial and environmental outcomes under cooperation mode than it gives under stand-alone mode. The authors observe that level of carbon price (used while estimating the carbon emissions) and ethanol tax credit do not effect the location and biomass processing capacity of biorefineries. However, both of these factors have a huge impact on the selection of biomass conversion technology.

Sharma et al. [48] formulate a scenario optimization model that considers weather uncertainty in biomass supply chain, and conduct a case study at a biorefinery in Kansas. The model objective function minimizes the annual cost of biomass supply to biorefineries including procuring, harvesting, transportation and storage costs. The model determines the acres of land leased for biomass production, the material flow from farms to biorefineries and the number of machinery units (for harvesting and transportation) rented or purchased. This model captures the influence of weather uncertainty on purchasing and deployment of assets. The authors test this model according to 12 different weather scenarios each having equal probability to occur at

a biorefinery in Kansas. Their case study results show that available harvest work hours affect the major cost related decisions in the biomass supply chain. Moreover, biomass yield is shown to be a significant factor while designing the supply chain.

Tay et al. [51] present a robust optimization approach that incorporates uncertainties in raw material (biomass) supply and product (biofuel) demand. A non-linear mixed-integer program is formulated with multiple biomass types and technology pathways to convert these biomass types to different biofuels demanded. The objective of the model is to determine the optimal integrated biorefinery configuration while maximizing the average net present revenue of the biorefinery. The proposed model is solved using off-the-shelf software (LINGO 10.0 with Global Solver). A case study that considers a single biomass source i.e., black liquor (BL) and 8 alternative pathways to process BL is conducted. A total of 4 scenarios with different potential biomass supply and biofuel demand are considered.

3.3 Problem Definition and Formulation

We consider a three level (i) farm, (ii) collection facility and (iii) biorefinery, biomass supply chain network depicted in Figure 3.1. The objective is to minimize the expected total system cost while satisfying certain problem constraints.

Farms are the supply points where biomass is produced and their locations are known. We assume the biomass supply quantity at farms depends on two factors: (i) size (acreage) of biomass planting area and (ii) biomass yield. Biofuel producer offers a unit biomass price to a set of farmers and makes a commitment of buying all the biomass supply they produce. In fact, this is similar to a wholesale price contract between the farmers and the biofuel producer. Based on this commitment, biofuel producer purchases all the supply produced at farms. Farmers determine the size of their biomass planting area based on the wholesale price offered to them and

as this wholesale price increases biomass plating acreage likely to get larger. After harvested at farms, biomass is shipped to biorefineries through collection facilities. All biomass shipment is done through collection facilities and direct shipments from farms to biorefineries are not allowed.

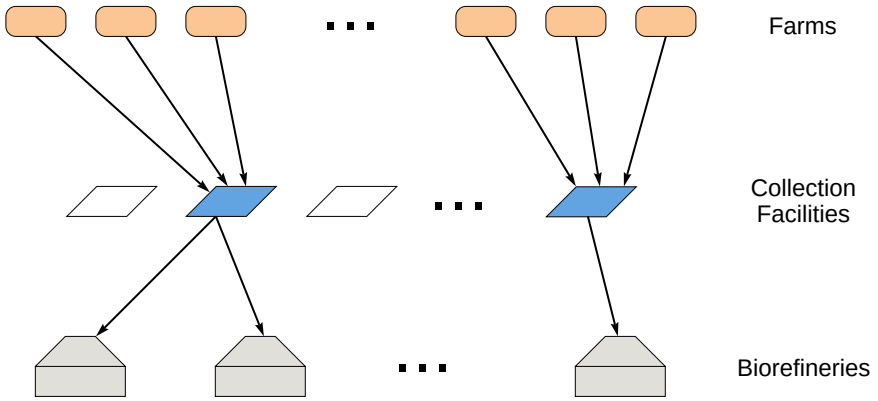


Figure 3.1: Biomass supply chain structure

Collection facilities are hub locations where biomass coming from different farms is consolidated and shipped to biorefineries. The main incentive to use collection facilities is the transportation economies-of-scale they offer. Collection facilities provide transportation cost savings since shipments are made in large quantities and possibly with different and more efficient modes of transportation. Collection facility locations and capacity levels are to be determined by the biofuel producer and the locations are chosen among a set of known candidate points.

Biorefineries demand biomass to process and convert it to biofuel. We assume that biorefinery locations and their biomass demands are deterministic and known. If demand at biorefineries is not satisfied with the biomass coming from farms, it

is out-sourced from an outside supplier via collection facilities, and a penalty cost is incurred for each unit of out-sourced biomass. On the other hand, if the biofuel producer has surplus biomass after satisfying all the demand, it salvages all the excess amount.

We model this problem as a two-stage stochastic program based on scenarios. In the first-stage, the collection facility locations and sizes, as well as the biomass wholesale price offered to the farmers are determined. These decisions, also refereed as here-and-now decisions, are taken prior to knowing the biomass yield for each farm.

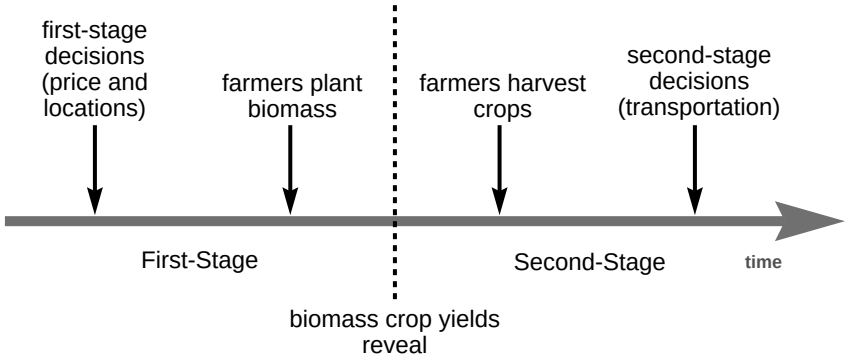


Figure 3.2: Time-line of events

Figure 3.2 summarizes the time-line of events in this system. First, the first-stage decisions (collection facility location and capacity levels and biomass wholesale price) are determined. Second, according to the offered wholesale price, each farmer decides how much land (acreage) to allocate for biomass production. The biomass quantity (supply) at each farm depends on both the size of the planted biomass land, and the biomass yield, which is affected by random events such as weather and

other environmental factors. Hence, the uncertainty of each farm's biomass yield is addressed via a finite set of scenarios. One of these scenarios happen and the biomass yield for each farm is revealed. In the second-stage, farmers harvest all the biomass they produce. At this point, the supply at each farm is known and the second-stage decisions are taken by the biofuel producer based on these supply amounts. Following decisions are determined in the second-stage: (i) shipment quantities from farms to biorefineries through collection facilities, (ii) out-sourced biomass quantities at each biorefinery, and (iii) salvaged biomass quantities at each farm. To formulate this problem, we introduce the notation as follows:

Sets:

$\mathcal{I}$ set of farms indexed by i

$\mathcal{J}$ set of collection facilities indexed by j

$\mathcal{K}$ set of biorefineries indexed by k

$\mathcal{L}$ set of capacity (size) levels indexed by l

$\mathcal{S}$ set of yield scenarios indexed by s

Random Variable:

Φ_i random variable for biomass yield per acre at farm i

Parameters:

$\Gamma_i(\pi)$	function that gives the acre of biomass planted in farm i according to wholesale price π
Φ_i^s	biomass yield per acre at farm i according to scenario s, a realization of Φ_i
f_{jl}	fixed cost of opening and operating a collection facility j of size l
c	unit penalty cost of out-sourced biomass
h	unit salvage cost of biomass
p^s	probability of occurrence of scenario s
k_l	capacity of size l collection facility
d_k	quantity of biomass demanded at biorefinery k
t_{ijk}	per unit transportation cost from farm i to biorefinery k through collection facility j

Decision Variables:

Y_{jl}	takes value 1, if a collection facility of size l is opened at location j and takes value 0, otherwise
π	unit biomass wholesale price
X_{ijk}^s	quantity of biomass shipped from farm i to biorefinery k through collection facility j for scenario s
Q_k^s	quantity of out-sourced biomass at biorefinery k for scenario s
W_i^s	quantity of biomass salvaged to farm i for scenario s.

According to the notation introduced, the two-stage stochastic program is formulated as follows:

$$\text{Min} \quad \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} f_{jl}\, Y_{jl} + \mathbb{E}\big[\Psi(Y_{jl}, \pi, \Phi_i)\big] \tag{3.1}$$

subject to

$$\sum_{l\in\mathcal{L}} Y_{jl} \leq 1 \qquad \forall\, j \in \mathcal{J} \tag{3.2}$$

$$\sum_{j\in\mathcal{J}} \sum_{l\in\mathcal{L}} k_l\, Y_{jl} \geq \sum_{k\in\mathcal{K}} d_k \tag{3.3}$$

$$h \leq \pi \leq c \tag{3.4}$$

$$Y_{jl} \in \{0,1\} \qquad \forall\, j \in \mathcal{J},\, l \in \mathcal{L} \tag{3.5}$$

where for a given particular realization, i.e., scenario s,

$$\Psi(Y_{jl}, \pi, \Phi_i^s) = \text{Min} \quad \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} t_{ijk}\, X_{ijk}^s + \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{I}} \Phi_i^s\, \pi\, \Gamma_i(\pi) + \sum_{k\in\mathcal{K}} c\, Q_k^s - \sum_{i\in\mathcal{I}} h\, W_i^s \tag{3.6}$$

subject to

$$\sum_{i\in\mathcal{I}} \sum_{k\in\mathcal{K}} X_{ijk}^s \leq \sum_{l\in\mathcal{L}} k_l\, Y_{jl} \qquad \forall\ j \in \mathcal{J} \tag{3.7}$$

$$\sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} X_{ijk}^s + W_i^s = \Gamma_i(\pi)\, \Phi_i^s \qquad \forall\, i \in \mathcal{I} \tag{3.8}$$

$$\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} X_{ijk}^s + Q_k^s \geq d_k \qquad \forall\, k \in \mathcal{K} \tag{3.9}$$

$$X_{ijk}^s,\, W_i^s,\, Q_k^s \geq 0. \tag{3.10}$$

Objective function (3.1) includes the fix cost and the expected value of (3.6) over the set of scenarios $\mathcal{S}$. Constraint (3.2) forces at most one collection facility with a capacity to open at each candidate location. Since all biomass transportation (including out-sourced biomass from outside suppliers) goes through the opened collection facilities, the system must have a total capacity larger than the total sys-

tem demand. Therefore, constraint (3.3) is a surrogate constraint, which sets the minimum collection facility capacity requirement for the entire system. The biomass wholesale price must be greater than the salvage price and also it must be less than the penalty (out-sourcing) cost. If the biomass wholesale price is less than the salvage price, then the biofuel producer makes profit just by salvaging biomass. Conversely, if the biomass wholesale price is greater than the unit penalty cost, then all demand is satisfied from outside suppliers. Hence, constraint (3.4) defines lower and upper bounds for unit biomass wholesale price.

Problem Ψ minimizes the total cost of transportation, purchasing and penalty, while also considering the total revenue coming from salvaging. Notice the last term of the objective function (3.6) denotes this revenue, hence, it has a negative sign. Constraint (3.7) is the capacity constraint on each collection facility. Constraint (3.8) is the supply constraint which ensures all supply to be transported to biorefineries or sold back to farmers. Note that right hand side of constraint (3.8) is different for each farm and for each scenario. Constraint (3.9) is to satisfy biomass demand at each biorefinery. Problem Ψ is updated for different realizations of the random variable Φ_i since the second term in the objective function (3.6) and the right-hand-side of constraint (3.8) change for each $s \in \mathcal{S}$.

This stochastic program has complete recourse. In other words, for every first-stage decision, problem Ψ is feasible for all scenarios. This is because of two reasons. First, constraint (3.3) ensures that there is enough capacity to transport biomass through collection facilities, thus, all the collection facility locations and capacities determined in the first-stage are feasible. The second reason is the option of satisfying demand with out-sourcing. Regardless of the biomass wholesale price and the supply availability at farms, the demand at each biorefinery can always be satisfied by out-sourcing.

The nature of the mathematical model presented above depends highly on $\Gamma_i(\pi)$. Therefore, in Section 3.4, we explain how the farmers' decision to plant biomass is affected by the biomass wholesale price, which determines the structure of the function Γ.

3.4 Farmer's Decision Model

In order to determine the Γ function introduced in the mathematical formulation, farmer's decision must be taken into consideration. In other words, we need a model to determine the relationship between biomass energy crop i.e., switchgrass, price and supply. There have been some studies on this topic in the EU countries. Bocqueho and Jacquet [8] look into the effect of farmers liquidity constraints and risk preferences on energy crop adoption by farmers. Their study considers a multi-period decision process for farmer. In this study, we follow a similar approach that is applied in Downing and Graham [17] in order to determine the relationship between switchgrass price and supply.

We assume that a farmer makes his planting decision based solely on his expected profit. Therefore, a farmer adopts a new crop, if the expected profit of the new crop is higher than the expected profit of at least one of the crops that the farmer currently plants. Let's assume farmer i plants a set of crops which is denoted by $\mathcal{R}_i$ and each crop is indexed by r. Moreover, we introduce the notation below to construct farmer's decision model:

YLD_{ir} the expected yield of crop r estimated by farmer i (unit/acre)

PRC_r the expected market price of crop r (\$/unit)

CST_{ir} the production cost of crop r by farmer i (\$/acre)

YLD_{i*} the expected yield of switchgrass by farmer i (dt/acre)

BRK_{ir} the break-even price of farmer i for crop r (\$/dt)

CST_{i*} the production cost of switchgrass by farmer i (\$/acre).

We know that for every farmer i and crop r the following equation must hold:

$$(YLD_{ir}\, PRC_r) - CST_{ir} = (YLD_{i*}\, BRK_{ir}) - CST_{i*}. \tag{3.11}$$

The left and the right-hand-sides of equation (3.11) represent the expected unit profit (\$/acre) that farmer i makes by planting crop r and switchgrass, respectively. Farmer i plants switchgrass only when his expected profit from it is greater than his expected profit from at least one of the crops in $\mathcal{R}_i$. Notice that YLD_{i*} represents the expected switchgrass yield for farmer i. This can be the average value calculated using all scenarios ($\sum_{s\in\mathcal{S}} \Phi_i^s/|\mathcal{S}|$)). This value is the expected value that farmer i uses when he makes his planting decision. Similarly, YLD_{ir} represents the expected yield for crop type r estimated by farmer i. CST_* and CST_r denote the expected cost for planting, producing and harvesting of switchgrass and crop r respectively. For farmer i, there are $|\mathcal{R}_i|$ many crops and hence, that many break-even prices, each represented by BRK_{ir} where $r \in \mathcal{R}_i$. These prices can be determined by solving (3.11) for BRK_{ir}, which is the break-even price that would convince farmer i to plant switchgrass instead of crop r. There may be different break-even prices for different crops and farmers. After solving equation (3.11) for each farmer $i \in \mathcal{I}$ and his crops $r \in \mathcal{R}_i$ and determining BRK_{ir}'s, we obtain the function $\Gamma_i(\pi)$ as a stepwise function.

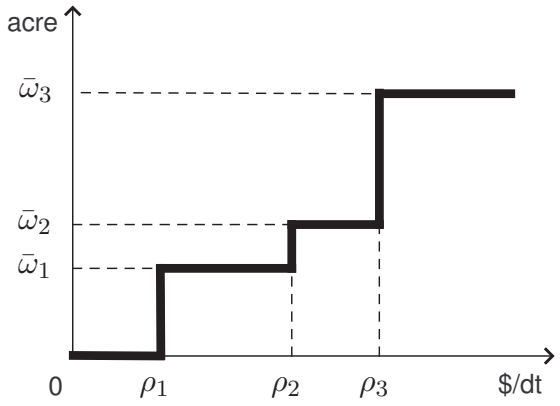


Figure 3.3: Stepwise Gamma (Γ) function

Figure 3.3 illustrates the shape of the gamma function for farmer i, i.e., $\Gamma_i(\pi)$. In this illustration, there are three crops that farmer i currently plants, i.e., $|\mathcal{R}_i| = 3$. The x-axis represents wholesale price (π) and ρ_{i1}, ρ_{i2} and ρ_{i3} represent the break-even prices for each crop. The y-axis shows how much acre farmer i plants according to given π. Notice that $\bar{\omega}_1$ represents the acreage that is currently planted for crop 1. Similarly, crop 2 and crop 3 occupy $(\bar{\omega}_2 - \bar{\omega}_1)$ and $(\bar{\omega}_3 - \bar{\omega}_2)$ acreages respectively. Now, we present a key property that helps us to reformulate the overall stochastic program presented earlier.

Proposition 3.4.1. *Optimal wholesale price offered to a farmer is one of his break-even prices.*

The proof is simple. Suppose the optimal price is not one of the break-even prices. This means, it can be written as $(\rho_{ir} + \epsilon)$ for some r and where $\epsilon \geq 0$. Without loss of generality, let's assume scenario s happens. This means that the biofuel producer pays farmer i a total of $((\rho_{ir} + \epsilon)\,\bar{\omega}_r\,\Phi_i^s)$ to purchase $(\bar{\omega}_r\,\Phi_i^s)$ units of biomass. However, he can always reduce the price to ρ_{ir} and pay $(\rho_{ir}\,\bar{\omega}_r\,\Phi_i^s)$ to purchase the same amount of biomass.

Corollary 3.4.2. *Optimal wholesale price offered to a group of farmers is one of the break-even prices of a farmer in the group.*

The proof is similar to the previous proof. The result of Proposition 3.4.1 can be generalized for a group of farmers.

Equation (3.11) uses farmer's yield estimates on his current crops and switchgrass. Depending on how the farmer estimates the expected yields for switchgrass and his current crops, the resulting price break-even points change. If a farmer is optimistic about switchgrass yield, he estimates a high switchgrass yield. This results in lower break-even prices for the farmer and he becomes more willing to switch his production to switchgrass. In other words, he is willing to take risks. Similarly, if a farmer is pessimistic about switchgrass yield, he estimates a low switchgrass yield. In that case, the break-even prices increase and the farmer becomes less interested in switching his production to switchgrass. We refer this kind of farmers as risk averse farmers. Note that different yield estimates does not change the structure of the stepwise function but they only changes the x-axis (break-even prices) of the function. The y-axis (planted acres) stays the same since we assume the farmer already allocates some acreage to each crop he currently plants. An analysis on this issue is done and the effects of farmer's decision model on the overall problem is presented in Section 3.9.3.

3.5 Reformulated Model

We know from Corollary 3.4.2 that the optimal price value is one of the break-even prices of a farmer. Therefore, we can reduce the set of continuous price values to a set of discrete price points, since there are a finite number of break-even prices for each farmer. Let's define set $\mathfrak{F}$ that represents all the break-even prices for all farmers ($\mathfrak{F} = \bigcup_{i,r} \rho_{ir} \ \forall i \in \mathcal{I}, r \in \mathcal{R}_i$). We know that some of these prices might be the same and might not satisfy constraint (3.4). Therefore, we define a new set

$\mathcal{F} \subseteq \mathfrak{F}$ that represents all the possible distinct prices. Moreover, let P_f where $f \in \mathcal{F}$, be a binary decision variable, which takes 1 if price f is chosen, and 0 otherwise, and let π_f be the associated price value. Furthermore, let ω_{if} denote the acreage of switchgrass planted by farmer i when price f is offered. Note that parameter ω_{if} is equal to $\Gamma_i(\pi_f)$, and can easily be obtained.

Recall that all the biomass produced by farmers is purchased by the biofuel producer. Therefore, for each scenario, purchasing cost is only a function of π_f and we know the probability of each scenario (p^s). Hence, purchasing cost term can be taken out of problem Ψ and can be placed separately in the objective function (3.1). According to these modifications, the model presented in Section 3.3 is reformulated as below:

$$\text{Min} \quad \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} f_{jl}\, Y_{jl} + \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}} p^s\, \Phi_i^s \omega_{if}\, \pi_f P_f + \mathbb{E}\big[\Delta(Y_{jl}, \pi, \Phi_i)\big] \tag{3.12}$$

$$\text{subject to} \;\; (3.2), (3.3), (3.5)$$

$$\sum_{f\in\mathcal{F}} P_f = 1 \tag{3.13}$$

$$P_f \in \{0,1\} \qquad \forall\, f \in \mathcal{F} \tag{3.14}$$

where for a given particular realization, i.e., scenario s,

$$\Delta(Y_{jl}, P_f, \Phi_i^s) = \text{Min} \quad \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} t_{ijk} X_{ijk}^s + \sum_{k\in\mathcal{K}} c\, Q_k^s - \sum_{i\in\mathcal{I}} h\, W_i^s \tag{3.15}$$

subject to $(3.7), (3.9), (3.10)$

$$\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} X^s_{ijk} + W^s_i = \sum_{f\in\mathcal{F}} \Phi^s_i \omega_{if} P_f \qquad \forall\, i \in \mathcal{I}. \tag{3.16}$$

In this new formulation, first and second terms of the objective function (3.12) denote the total fixed cost and the expected purchasing cost, respectively. Constraint (3.4) is taken out of the formulation since it is no longer necessary. Instead, constraint (3.13) is added to the formulation, which makes sure only one of the break-even prices is selected as price.

Notice that the overall problem becomes linear with this new formulation. Although this formulation introduces new binary variables into the model, i.e., P, and increases the number of decision variables, it also reduces the pricing decision to a finite set of choices.

3.6 Solution Methodology - L-Shaped Method

In essence, L-shaped method is the same as Benders Decomposition algorithm, which decomposes the overall mathematical problem into a master problem and a subproblem [23]. In L-shaped framework, first-stage problem (FSP) and second-stage problem (SSP) correspond to the master problem and subproblem, respectively. The dual of the second-stage problem (SSDP) is easily obtained and utilized to construct Benders type cuts, which are added to the FSP via the use of an auxiliary continuous variable. In each iteration of the L-shaped method, FSP provides a lower bound to the overall problem along with a set of first-stage decisions. These first-stage decisions are fixed in SSDP, whose solution provides a set of dual variable values that are used to construct a Benders type cut. SSDP solution together with the FSP solution is used to calculate an upper bound to the overall problem. These three steps (i) solving SSDP and generating Benders type cut using the SSDP solution, (ii) adding

the generated cut to the FSP and solving the new FSP and (iii) constructing a new SSDP based on the FSP solution, are continued to be executed until a desired optimality gap between the upper bound and lower bound is obtained. L-shaped method guarantees to converge to the optimal solution.

Flippo and Rinnooy Kan [20] states the conditions where Benders Decomposition algorithm, similarly the L-shaped method, can be employed to solve a mathematical problem. Those conditions are satisfied for our problem since (i) SSDP is a linear program and (ii) none of the dual constraint RHSs depend on the FSP variables. Therefore, we can utilize L-shaped method to solve this problem.

3.6.1 Second-Stage Dual Problem (SSDP)

We construct the SSDP by assuming given values of the first-stage integer variables i.e., $\hat{Y}_{jl}$, and $\hat{P}_f$. For each scenario s, we define the dual variables λ_j^s, β_k^s and α_i^s associated with constraints (3.7), (3.9) and (3.16), respectively. SSDP is formulated for each scenario $s \in \mathcal{S}$ as follows:

$$\text{Max} \quad \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} k_l \hat{Y}_{jl} \lambda_j^s + \sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}} \Phi_i^s \omega_{if} \hat{P}_f \alpha_i^s + \sum_{k\in\mathcal{K}} d_k \beta_k^s \tag{3.17}$$

subject to

$$\lambda_j^s + \alpha_i^s + \beta_k^s \leq t_{ijk} \qquad \forall\, i \in \mathcal{I},\, j \in \mathcal{J}\, k \in \mathcal{K} \tag{3.18}$$

$$\beta_k^s \leq c \qquad \forall\, k \in \mathcal{K} \tag{3.19}$$

$$\alpha_i^s \leq -h \qquad \forall\, i \in \mathcal{I} \tag{3.20}$$

$$\lambda \leq 0,\ \ \beta \geq 0,\ \ \alpha \text{ unrestricted.} \tag{3.21}$$

We let $\mathcal{E}^s$ denote the set of extreme points of the polyhedron defined by (3.18)-(3.21) for scenario s. Moreover, $\lambda_j^{e^s}$, $\alpha_i^{e^s}$ and $\beta_k^{e^s}$ and η^{e^s} denote the dual variable and objective function value for the extreme point $e^s \in \mathcal{E}^s$. Furthermore, let η^{*^s} and η^* be the optimal objective values for the SSP corresponding to scenario s and the overall SSP respectively. Note that $\eta^* = \sum_{s\in\mathcal{S}} p^s\, \eta^{*^s}$. Since $\eta^{e^s} \leq \eta^{*^s},\ \forall e^s \in \mathcal{E}^s$, for each scenario s, we can reformulate SSDP as $\min_{\eta^s \geq 0}\{\eta^s : \eta^{e^s} \leq \eta^s,\ \forall e^s \in \mathcal{E}^s\}$ where

$$\eta^{e^s} = \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} k_l\, \hat{Y}_{jl}\, \lambda_j^{e^s} + \sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}} \Phi_i^s\, \omega_{if} \hat{P}_f\, \alpha_i^{e^s} + \sum_{k\in\mathcal{K}} d_k\, \beta_k^{e^s}. \tag{3.22}$$

Moreover, considering the expectation of all scenarios, the overall SSDP can be restated as $\min_{\eta\geq 0}\{\eta : \sum_{s\in\mathcal{S}} p^s\, \eta^{e^s} \leq \eta,\ \forall e^s \in \mathcal{E}^s\}$ or equally, $\min_{\eta\geq 0}\{\eta : \eta^e \leq \eta,\ \forall e \in \mathcal{E}\}$ where

$$\eta^e = \sum_{s\in\mathcal{S}} p^s \left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} k_l\, \hat{Y}_{jl}\, \lambda_j^{e^s} + \sum_{i\in\mathcal{I}} \Phi_i^s\, \omega_{if} \hat{P}_f\, \alpha_i^{e^s} + \sum_{k\in\mathcal{K}} d_k\, \beta_k^{e^s} \right). \tag{3.23}$$

3.6.2 Reformulation of the First-Stage Problem (FSP)

Below, we present a reformulation of the original two-stage stochastic problem using the above representation of SSDP.

$$\text{Min} \quad \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}} f_{jl}\, Y_{jl} + \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}} p^s\, \Phi_i^s \omega_{if}\, \pi_f P_f + \eta \tag{3.24}$$

subject to (3.2), (3.3), (3.5), (3.13),

$$\eta^e \leq \eta \qquad \forall\, e \in \mathcal{E} \tag{3.25}$$

$$\eta \geq 0. \tag{3.26}$$

Constraint (3.25) is written for all the extreme points in SSDP polyhedron i.e., set $\mathcal{E}$. These extreme points are not available beforehand and will be iteratively added to the FSP in the L-shaped framework. Hence, a relaxed FSP is solved in every iteration where constraint (3.25) is written only for a subset of $\mathcal{E}$.

Notice that the auxiliary variable η represents the expectation of SSP by considering all scenarios and generating a single Benders cut as in the form of (3.23). This is referred as the single-cut approach in the L-shaped method context. Alternatively, a cut for each scenario can be generated using a set of auxiliary variables, corresponding to each scenario s as in the form of (3.22). In that case, we have $|\mathcal{S}|$ many auxiliary variables i.e., η^s, and the third term in (3.24) is modified as $\sum_{s\in\mathcal{S}} \eta^s$, so that it considers all scenarios. This is referred as the multi-cut approach. Another alternative between single-cut and multi-cut can be taken in which some of the scenarios are combined and multiple-cuts are generated. This combination can be done in many different ways, and it is up to the user to decide how to combine the scenarios. We will explain our particular scenario combination approach in this context in Section 3.6.4.

3.6.3 Strengthened Benders Cuts

Traditional Benders cuts might not perform well in most cases and further improvement might be required to achieve higher performance. Magnanti and Wong [34] define the strongness of a cut for an optimization problem $\min_{y\in\mathcal{Y},z\in\mathcal{R}}\{z\colon f(u)+yg(u)\leq z, \forall u\in\mathcal{U}\}$ as follows: The cut $f(u_1)+yg(u_1)\leq z$ dominates or is stronger than another cut $f(u_2)+yg(u_2)\leq z$, if $f(u_1)+yg(u_1)\geq f(u_2)+yg(u_2),\ \forall y\in\mathcal{Y}$ with a strict inequality for at least one $y\in\mathcal{Y}$.

The strength of the cut (3.22) depends on the values of $\hat{\alpha}_i$, $\hat{\beta}_k$ and $\hat{\lambda}_j$ i.e., the optimal solution obtained in SSDP. Notice that $\hat{\alpha}_i$ and $\hat{\beta}_k$ are fixed and can not be

changed for given first-stage decisions while maintaining SSDP optimality. However, λ_j's can be improved for those j's that are not selected to open a facility in the first-stage. Let's define set $\hat{\mathcal{J}}$ as the set of locations selected to open collection facilities in the first-stage. The first term in SSDP objective function (3.17) depends on the first-stage location and capacity decisions. Notice that the cost coefficient of λ_j is $\sum_{l\in\mathcal{L}} k_l$ for $j \in \hat{\mathcal{J}}$, and it is zero for $j \in \mathcal{J} \setminus \hat{\mathcal{J}}$. This means that while solving the SSDP, λ_j variables become irrelevant (redundant) for j's that are not selected as facility locations, i.e., $j : \sum_{l\in\mathcal{L}} \hat{Y}_{jl} = 0$. Although the value of those λ_j's do not have an effect on the SSDP objective function value, they have an effect on the quality of the Benders cut generated. In other words, alternate optimal solutions might exist in the SSDP problem, which result in different Benders type cuts. Therefore, the best values for λ_j's need to be selected for the Benders cut to be the strongest. We must solve an auxiliary problem to find the best SSDP solution which yields the best cut. This auxiliary problem aims to find the maximum λ_j values while maintaining the SSDP optimality. Constraint (3.18) is the only constraint in SSDP that restricts λ_j's and for fixed α_i and β_k values, and also considering the fact that λ_j's are non-positive, the maximum λ_j's can be determined using equation (3.27) below:

$$\lambda_j = \min\{\min_{i,k}\{t_{ijk} - \hat{\alpha}_i - \hat{\beta}_k\}, 0\} \qquad \forall j \in \mathcal{J} \setminus \hat{\mathcal{J}}. \tag{3.27}$$

We use equation (3.27) to determine the updated λ_j values for $j \in \mathcal{J} \setminus \hat{\mathcal{J}}$, and use $\hat{\lambda}_j$ values coming from the SSDP optimal solution for $j \in \hat{\mathcal{J}}$. This set of new λ_j values are the best possible to generate the strongest Benders type cut. We use the same α_i and β_k values coming from SSDP optimal solution, i.e., $\hat{\alpha}_i$ and $\hat{\beta}_k$, to generate the cuts.

3.6.4 Scenario Generation and Multiple-cuts

The supply points (farmers) can be grouped based on their geographical locations. Let's assume there are m farmer groups (sets) denoted by $\mathcal{I}_g$ where $\bigcup_{g=1}^{m} \mathcal{I}_g = \mathcal{I}$. Moreover, we assume that there are two (high and low) switchgrass yield levels for each farm group g. Hence, there are total of $n = 2^m$ different combinations, which we call "scenario types". We can partition the whole scenario set $\mathcal{S}$ into n subsets where each subset $\mathcal{S}_t$ contains the scenarios of type t. This way of scenario generation intuitively makes sense since the farmers that are close to each other are likely to be affected by similar weather and environmental conditions. Moreover, we can argue that the soil of the farms that are close to each other have similar characteristics and this also results in similar switchgrass yields. While generating the yield levels for each farm group g, we use $U[a_g, b_g]$ for low level yields and $U[b_g, c_g]$ for high level yields where $a_g \leq b_g \leq c_g, \ \forall g = 1 \dots m$. In other words, the switchgrass yield for farm $i \in \mathcal{I}_g$ in scenario $s \in \mathcal{S}_t$, i.e., Φ_i^s, is generated using $U[a_g, b_g]$ or $U[b_g, c_g]$ based on the scenario type t, which is obtained by generating a random integer variable in $[1, n]$.

Following a similar approach in Section 3.6.1, SSDP can be reformulated as $\min_{\eta^t \geq 0}\{\eta^t : \sum_{s \in \mathcal{S}_t} p^s \, \eta^{e^s} \leq \eta^t, \ \forall e^s \in \mathcal{E}^s\}$ where η^{e^s} is defined as in (3.22). In other words, instead of generating a single Benders cut in every iteration by aggregating all the scenarios in $\mathcal{S}$ as in (3.23), n number of Benders cuts are generated for each scenario type t by only aggregating scenarios in set $\mathcal{S}_t$. Notice that in this reformulation, there are n auxiliary variables, i.e., η^t, to represent the SSP objective function value. Therefore, the last term of FSP objective function (3.24) is modified

as $\sum_{t=1}^{n} \eta_t$. Moreover, constraint sets (3.25) and (3.26) are modified as:

$$\sum_{s \in \mathcal{S}_t} p^s \eta^{e^s} \leq \eta^t \qquad \forall\, t = 1, \ldots, n, \;\; e^s \in \mathcal{E}^s \tag{3.28}$$

$$\eta^t \geq 0. \tag{3.29}$$

As the number of farmer groups (m) increases, the number of scenario type (n) increases and therefore, we add more cuts in each iteration of the L-shape method. Hence, our solution algorithm based on L-shaped method, which utilizes multiple strengthened Benders type cuts in every iteration is described below in Algorithm 2.

Algorithm 2 Multiple-cut L-shaped Method

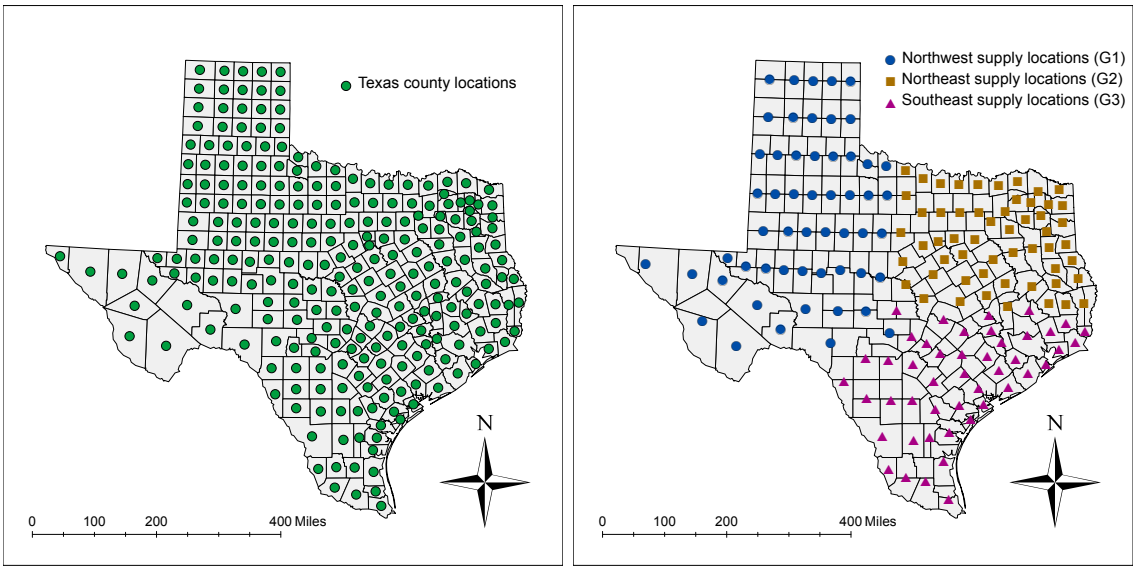
initialize $\epsilon = 0.02$, $\texttt{optgap} = 1.0$, $\texttt{LB} = -\infty$, $\texttt{UB} = \infty$, $\texttt{UB}^{\texttt{best}} = \infty$
while $(\texttt{optgap} > \epsilon)$ **do**
  Solve **FSP** and obtain the optimum objective value Z_{FSP} and its solution vectors $\hat{\mathbf{Y}}$ and $\hat{\mathbf{P}}$
  Set $\texttt{LB} = Z_{FSP}$
  for All scenarios $s \in \mathcal{S}$ **do**
    Substitute $\hat{Y}_{jl}$ and $\hat{P}_f$ and solve **SSDP**
    Store the optimal objective value Z^s_{SSDP} and associated dual variable solutions $\hat{\alpha}^s_i$ and $\hat{\beta}^s_k$
    Update $\hat{\lambda}^s_j$ values according to (3.27)
  end for
  Generate n Benders type cuts as in (3.28) and add it to **FSP**
  Set $\texttt{UB} = Z_{FSP} - \sum_{t=1}^{n} \eta_t + \sum_{s \in \mathcal{S}} p^s Z^s_{SSDP}$
  if $\texttt{UB} < \texttt{UB}^{\texttt{best}}$ **then**
    $\texttt{UB}^{\texttt{best}} = \texttt{UB}$
  end if
  $\texttt{optgap} = (\texttt{UB}^{\texttt{best}} - \texttt{LB})/\texttt{LB}$
end while
return $\texttt{UB}^{\texttt{best}}$ and its corresponding solution vectors $\hat{\mathbf{Y}}$ and $\hat{\mathbf{P}}$

3.7 Data Gathering & Problem Parameters

In this section, we explain how the input data is obtained and utilized via Geographical Information Systems (GIS). We use the literature to obtain realistic data, but some parameters are approximated due to lack of information.

3.7.1 Switchgrass Supply & Farms

We test our model on a realistic data generated in the state of Texas. We determine the geographical center point of each county in Texas (total of 254) using GIS. Some of these counties are paired together by locating the mid-point between them to create a smaller and more representative set of supply points. Finally, we determine 155 supply points, i.e., farms. Figure 3.4a and Figure 3.4b show the original 254 counties and the selected 155 supply points and their associated farm groups, respectively.



(a) Geographical center points of all Texas counties

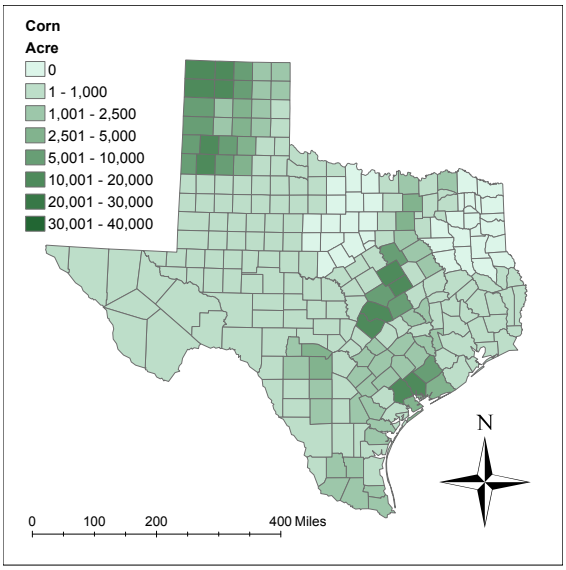
(b) Selected supply points and their associated groups

Figure 3.4: Biomass supply locations

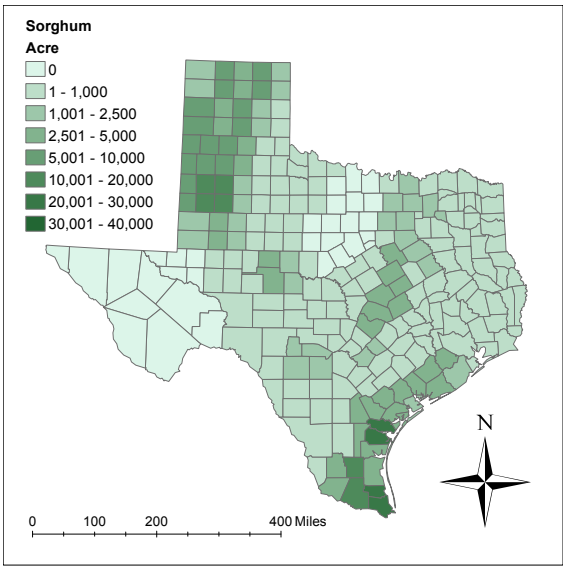
Supply is determined according to the farmer's decision model presented in Section 3.4. First, we select the four major crops planted in Texas in 2012 and 2013, which are corn, sorghum, cotton and wheat, to represent the crops that farmers currently plant. We use 2012 and 2013 data from United States Department of Agriculture [54] to determine the acre planted for each of these crops in each Texas county. We take 10% of these number to account for the participation (utilization) of farmers. These numbers correspond to the step sizes (ω) explained on Figure 3.3 in Section 3.4. Moreover, to determine the yield for each of these crops (YLD_{ir}), we again use the data provided by United States Department of Agriculture [54]. As the data indicates, each county has its own unique yield for each crop. For both the crop acreage and the crop yield, we take the average of the values in 2012 and 2013. Figure 3.5 shows the amount of planted acreage for the major crops in Texas. As we can see North Texas is where most of these crops are planted. In addition, there is a large amount of sorghum plantation in South Texas.

To determine the market prices for these four crops (PRC_r), we utilize the data provided by United States Department of Agriculture [58]. We use the price values for the state of Texas and make the appropriate conversions to determine the dollar per dry ton ($/dt) values. To estimate the production cost for these crops (CST_{ir}), we use the data provided by Texas A&M Agrilife Extension Service [52]. We take the average of the values in 2012 and 2013 to determine the market price and production cost of the selected crops. For each crop, we assume that production cost and market price are the same for all farms.

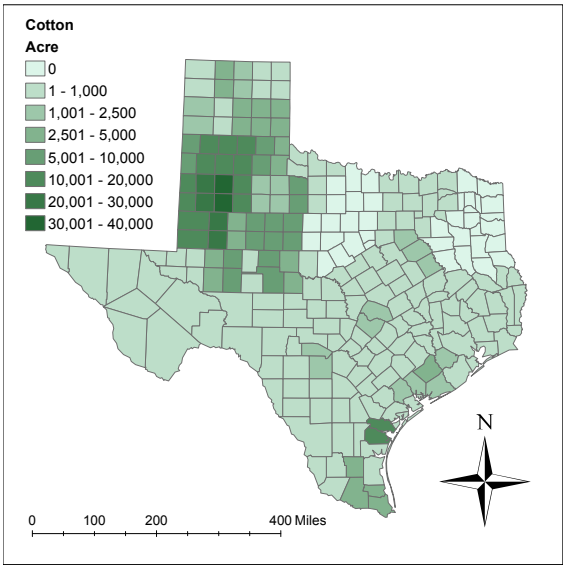
We assume the expected yield of switchgrass (YLD_{i*}) to be 6 dt/acre for all farms. Moreover, the production cost of switchgrass (CST_{i*}) is assumed to be the same for each county and estimated as $550/acre. We substitute the necessary parameters into equation (3.11) and solve it to find the switchgrass break-even prices for each



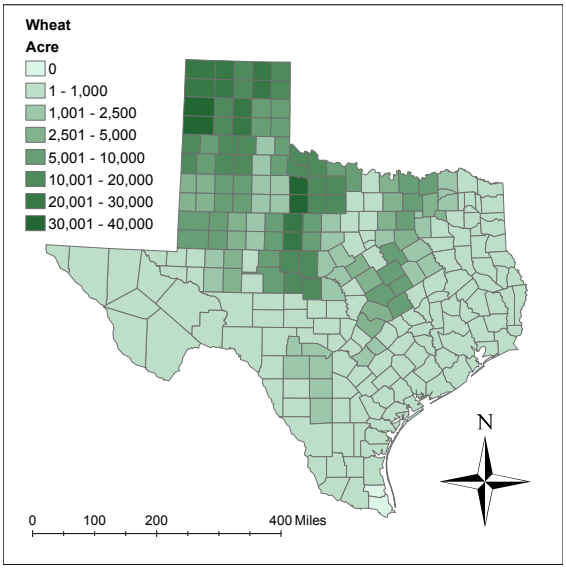
(a) Corn acre



(b) Sorghum acre



(c) Cotton acre



(d) Wheat acre

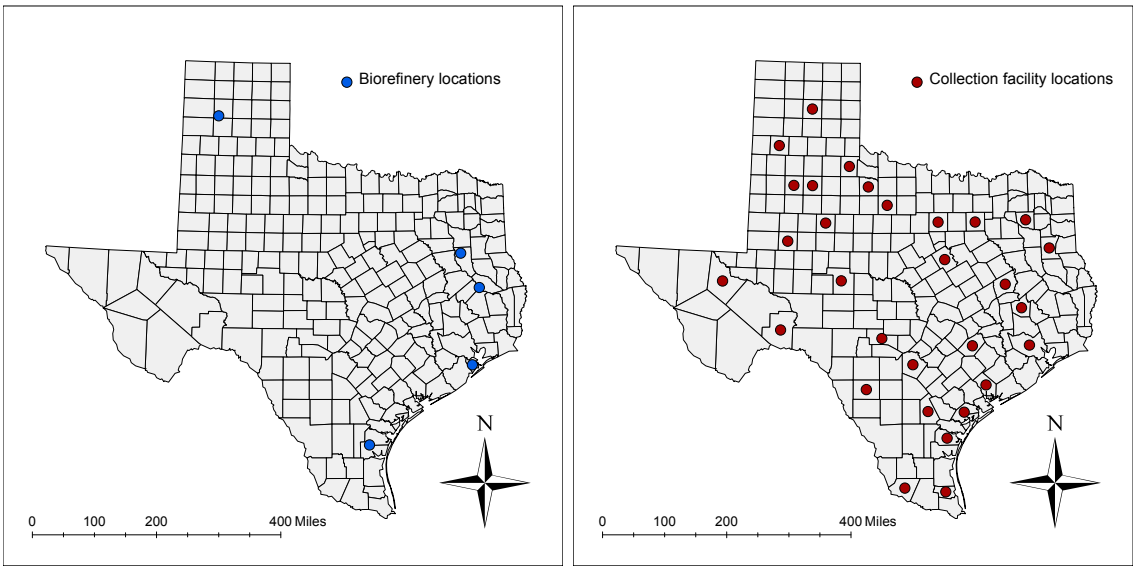
Figure 3.5: Acreage for different crops

farmer (BRK_{ir}). These numbers correspond to the break-even prices (π_f) in our reformulated model presented in Section 3.5.

3.7.2 Switchgrass Demand & Candidate Facility Locations

Figure 3.6 shows the candidate collection facility locations and the biorefinery locations (demand points). Five biorefineries, as shown in Figure 3.6a, are selected among the potential locations provided by United States Environmental Protection Agency [67]. In order to determine the total switchgrass demand, we use the estimated ethanol demand calculated using the Annual Refinery Report of EIA [63]. In this report 18 petroleum refineries with positive gasoline production are reported and using 50% utilization with E-10 (10% Ethanol) gasoline-ethanol production, we obtain around 710 million gallons of annual ethanol demand, as we also calculated in Section 2.6.3. We assume 33% of this total ethanol demand is satisfied converting switchgrass to biofuel. Therefore, the total demand for switchgrass is around 240 million gallons of ethanol equivalent switchgrass. We assume the switchgrass conversion rate to be 80 gallons/dt and this number is used to perform reverse conversion to calculate the total switchgrass amount demanded, which turns out to be around 3 million dt annually. Finally, this total demand is distributed equally to the selected biorefineries, each demanding 600,000 dt switchgrass.

Figure 3.6b shows the candidate collection facility locations. United States Department of Agriculture [55] divides Texas into 15 agricultural districts for statistical purposes. We randomly select two counties from each district and determine its geographical center point. These 30 points are selected as the candidate collection facilities used in our case study. Moreover, we consider two size levels (small and large) for collection facility capacity decisions. The capacity of small and large size collection facilities are assumed to be 500,000 dt and 1,000,000 dt, respectively.



(a) Biorefinery locations (b) Candidate collection facility locations

Figure 3.6: Biorefinery and candidate collection facility locations

Generating scenarios is an important aspect of stochastic programming. One of the important issues while generating our scenarios is to cover the whole sample space. We use three farm groups represented by G_1, G_2 and G_3 as also shown in Figure 3.4b. For each farm in each farm group we use U[2,6] for low yield level (L) and U[6,10] for high yield level (H). Table 3.1 shows 8 different scenario types created for our case study that cover all possible outcomes.

Table 3.1: Scenario types

Scenario	Farm Groups			Scenario	Farm Groups		
Type (t)	G1	G2	G3	Type (t)	G1	G2	G3
1	H	H	H	5	L	H	H
2	H	H	L	6	L	H	L
3	H	L	H	7	L	L	H
4	H	L	L	8	L	L	L

3.8 Computational Study

Using the data obtained from Section 3.7 first, we perform a runtime study on different cut generation approaches. Second, we perform an SAA analysis to statistically justify the closeness of our solutions to the true optimal solution. Lastly, we show the value of solving the stochastic programming problem over the expected value problem.

3.8.1 Single-cut Approach vs. Multiple-cut Approach

We generate 30 random instances for each scenario number and each cut approach to compare the solution times of our L-shaped method using single-cut and multiple-cut approaches. Every instance is solved until 2% optimality gap is reached, and the solution times are recorded.

Table 3.2: Runtimes of single-cut and multiple-cut approaches with 2% optimality gap stopping criterion

# of	Runtimes (secs)						Improvement
	Single-cut			Multiple-cut			
Scenarios	Ave	Max	Min	Ave	Max	Min	Ave(%)
40	1409	3496	692	959	1519	532	32
80	1872	2903	1282	1483	2471	809	21
120	2674	4707	1873	1736	3141	930	35
160	2941	4053	1874	2103	4117	1233	29
200	3570	8194	2074	2432	3732	1900	32

Table 3.2 shows the runtimes of different scenario size problems (40, 80, 120, 160 and 200) solved with single-cut and multiple-cut approaches. We provide the average (Ave), maximum (Max) and minimum (Min) of the solution times for all

scenario sizes considered. Moreover, the average runtime improvement percentages are presented in the last column of Table 3.2.

For all number of scenario sizes, we observe at least 21% average runtime improvement when multiple-cut approach is implemented. For each scenario size, the minimum runtimes of multiple-cut approach are smaller than the minimum runtimes of single-cut approach. Moreover, except for scenario size of 160, the maximum runtimes of multiple-cut approach are smaller than the minimum runtimes of single-cut approach. These results clearly show that multiple-cut approach performs better than the single-cut approach in terms of computational effectiveness. Hence, we implement multiple-cut approach while utilizing L-shaped method in our analysis provided in the following sections.

3.8.2 SAA Analysis

It is difficult and often times impossible to solve stochastic programs with significantly large number of scenarios. Therefore, methods like Sample Average Approximation (SAA) are implemented to find close optimal solutions. SAA is a Monte Carlo simulation based approach to solve large-scale stochastic programs [30]. The main idea of SAA is to solve the SP with a subset of scenarios several times and infer statistical information from these solutions. Solving the SP with a subset of scenarios gives a lower bound estimate on the true optimal objective value. This lower bound problem is solved several times with different set of scenarios and an average lower bound value is obtained. Using Central Limit Theorem (CLT), we can approximate this lower bound value data to normal distribution and hence, we can determine a confidence interval for the lower bound. A similar approach can be utilized to determine a confidence interval for the upper bound. A feasible first-stage solution, which we take the one that gives the lowest lower bound result, is fixed

and solved with a very large set of scenarios. The result provides an upper bound estimate on the true optimal. If this upper bound problem is solved a number of times, a confidence interval can be obtained using CLT. In order to obtain a $(1-\alpha)$ confidence intervals for lower and upper bounds, equation (3.30) is used where μ and σ values denote the average (mean) and standard deviation, respectively. Moreover, m is the batch size and represents the number of times the upper and lower bound problems are solved. Lastly, $z_{\alpha/2}$ is the z-value for $\alpha/2$.

$$\mathrm{CI}_b = \left[\mu_b - z_{\alpha/2}\frac{\sigma_b}{\sqrt{m_b}},\ \mu_b + z_{\alpha/2}\frac{\sigma_b}{\sqrt{m_b}}\right] \qquad b = \mathrm{LB}, \mathrm{UB} \tag{3.30}$$

After obtaining confidence intervals for the lower and upper bounds, we can justify the goodness of our upper bound solution by comparing the lower and upper bound confidence intervals. If the lower and upper bound confidence intervals are close enough, we can conclude that the solution at hand, i.e., the best upper bound solution, is close enough to the true optimal. If that's not the case, either the number of scenarios or the batch sizes of the lower and upper bound problems are increased and the corresponding confidence intervals are calculated again. More information on SAA and it's convergence to the true optimal solution can be found in Linderoth et al. [32].

In our SAA analysis, we use different batch sizes (10 and 20) and scenario numbers (40, 80, 120, 160 and 200). We construct a 95% confidence interval, i.e., $\alpha = 0.05$, for both LB and UB and observe how close these confidence intervals are to each other. To obtain a quality solution, the upper and lower bound averages must be close to each other, as well as the deviations from the mean should be small. In other words, lower and upper bound confidence intervals must intersect as much as possible.

Table 3.3: SAA results for different batch sizes and scenario numbers

Batch size (m)		Scenario #		Ave (μ)		Std (σ)		Gap (%)
LB	UB	LB	UB	LB	UB	LB	UB	CI
10	10	40	800	337,299,000	339,204,800	3,002,432	1,358,172	1.37
10	10	80	1600	336,368,800	337,876,700	1,206,844	425,031	0.75
10	10	120	2400	337,756,500	337,753,500	1,374,237	217,209	0.51
10	10	160	3200	337,140,800	337,706,800	1,354,137	326,582	0.50
10	10	200	4000	337,335,700	337,884,700	998,095	216,784	0.39
20	10	40	800	336,882,900	339,284,100	2,987,470	682,897	1.23
20	10	80	1600	336,747,500	337,176,200	1,571,314	380,704	0.41
20	10	120	2400	337,375,800	336,915,300	1,199,779	231,898	0.34
20	10	160	3200	337,311,650	337,115,000	1,252,069	402,715	0.33
20	10	200	4000	337,130,500	337,130,400	1,112,879	188,654	0.29

Table 3.3 shows the mean objective values (μ) and the standard deviations (σ) for different batch sizes (m) and scenario numbers for lower bound (LB) and upper bound (UB) problems. Moreover, we also provide the gap between the lower and upper bound confidence intervals (CIs). These gap values represent the difference between the maximum and minimum boundaries of LB and UB confidence intervals. This means that for small gap values the solution quality is better.

We observe that the mean values for LB and UB are close to each other for all scenario numbers, except 40. For most instances, as the number of scenarios increase, we observe smaller deviations from the mean. However, this comes with a cost of increase in computational time. Therefore, to find a good solution in a reasonable time, we decide to use 80 scenarios with batch sizes of 10. This gives us 95% confidence intervals that are 0.75% apart from each other, which is reasonably small to statistically justify the quality of the solution. Moreover, the standard deviations of UB are significantly smaller than those of LB. This means that once the first-stage decisions are determined, the resulting system is robust and the expected

total system cost does not deviate much. In all our consequent analysis in this section, we use the scenario size of 80 with batch sizes of 10 instances to obtain the solution.

3.8.3 The Value of Stochastic Programming

One might think that solving the stochastic program is difficult and try using an alternative approach to find a feasible solution. A natural alternative way is to solve a much simpler problem, in which all the random variables are replaced with their expected values. This problem is called the expected value problem, and let us denote its objective function value as EV. The optimal first-stage solution to the expected value problem is called the expected value solution. If we fix the expected value solution and solve the stochastic program with a very large set of scenarios, we obtain the expected result of using the expected value solution, EEV. This value represents how good or bad the expected value solution actually performs in reality. Then, the value of the stochastic solution (VSS) is defined as the difference between the expected result of using the expected value solution and the objective value determined by solving the stochastic problem with recourse denoted as SS [7]. In other words,

$$VSS = EEV - SS. \tag{3.31}$$

To find the expected value solution, we solve the two-stage SP problem defined by (3.1)-(3.10) considering only one scenario, i.e., the expected yield scenario. In the expected yield scenario, switchgrass yields are assumed to be at their expected values (6dt per acre) for all farms. After the problem is solved according to the expected yield scenario, the first-stage decisions, i.e., expected value solution, are recorded. In our result, the expected value solution of price is found as \$42.99/dt with a set

of opened collection facilities. Notice that this price is lower than the optimal price obtained solving the original SP. Then, we fix the expected value solution and solve the SP with 8,000 scenarios. We obtain the resulting objective function value, which is the EEV, around \$361 million. Recall that our stochastic solution yields objective values, which is the SS, around \$338 million. Therefore, we can obtain the value of stochastic solution as $VSS = \$361 - \$338 = \$23$ million. In other words, our stochastic solution performs 8% better than the expected value solution. This gap increases significantly when we change some of the problem parameters, especially when the penalty cost (c) increases. Our results show that if we increase the penalty cost to \$250/dt, this gap increases to 15% and when the penalty cost is \$300/dt, the gap reaches 23%. This is because the expected value solution can not provide enough supply to satisfy the demand in most scenarios, and as the penalty cost gets higher, the expected total system cost increases drastically. The stochastic solution on the other hand, considers even the worse case scenario and adjusts its first-stage decisions accordingly. Therefore, even if the penalty cost increases, the expected total system cost is not affected significantly.

3.9 Case Study

To demonstrate the capabilities of our model, we conduct a case study in Texas using the realistic data presented in Section 3.7. First, we provide results on the base setting where the nominal problem parameter values are used. Then, we alter some of the problem parameter values and observe their effects on the first-stage decisions. Lastly, we change parameter values in farmers' decision model to capture the effects of farmers' decision on the overall problem.

3.9.1 Base Setting Results

We use the nominal values, and obtain the acreage of switchgrass plantation in the supply locations, the optimal collection facility locations as well as capacities and the optimal price value, which is \$47.74/dt, shown in Figure 3.7.

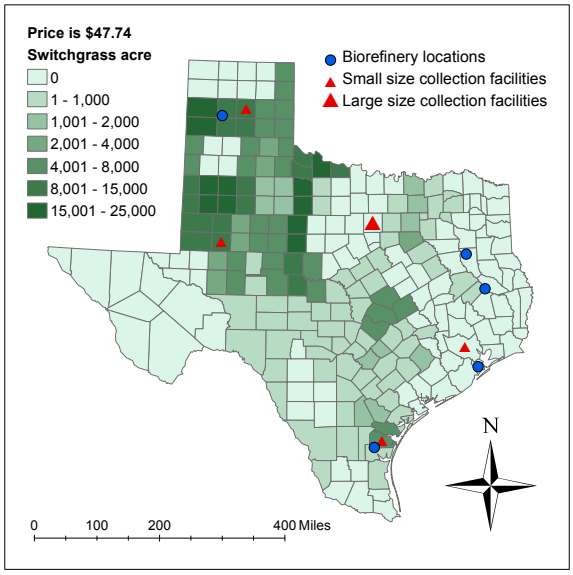


Figure 3.7: Base setting network analysis

The cropland acreage dedicated to switchgrass is shown in Figure 3.7 for all supply points, where darker colors represent larger switchgrass acreage. These values, i.e., ω_{if}, where f corresponds to the optimal price, are obtained using the farmers decision model presented in Section 3.4. Notice that the farms in North Texas plant the largest switchgrass areas based on the optimal price. However, this does not necessarily mean that the supply is highest in North Texas, since supply also depends on switchgrass yield, which is uncertain. Nevertheless, we can claim that the expected switchgrass supply is higher in North Texas compared to the other regions in Texas.

The optimal price ($47.74/dt) offered to the farmers provides around 585,000 acre of total switchgrass plantation in Texas. Assuming 6 dt/acre expected switchgrass yield results in 3.5 million dt expected total switchgrass supply in the system. Notice that this amount is more than the total demand, which is 3 million dt. In fact, in order to have around 3 million dt expected total switchgrass supply, the price should be set to $45.45/dt. Due to the penalty cost of out-sourced demand, the price is set higher than the expected value to consider even the worst case scenarios.

Moreover, five collection facilities, four in small size and one in large size, are opened around Texas as Figure 3.7 shows. Two of the small collection facilities located in North and Central Texas and the large collection facility in Central Texas serve the biorefineries in North and East Texas. The other two small collection facilities opened in South and East Texas serve the biorefineries in the same vicinity.

Our results show that the expected total system cost is around $338 million. More than half (around 53%) of the expected total system cost is for collection facility fix costs and expected switchgrass purchasing costs. Most of the remaining expected total system cost is for the transportation cost.

Notice that in the input data, four out of five biorefineries are located in East and South Texas where expected switchgrass supply is low. The optimal result indicates that transporting switchgrass from North Texas where the expected supply is high to East Texas where most of the demand occurs, is more economical compared to offering a higher price to the farmers and stimulate the farmers in East and South Texas to plant more switchgrass. This is the trade-off between the biomass wholesale price and the logistics costs in the system. Our results indicate that it is more economical to set a biomass wholesale price, which is not high enough for farmers in East and South Texas to switch production, but instead, transport the switchgrass supply in North Texas to satisfy the demand.

Fixed Wholesale Price

We fix the unit biomass wholesale price offered to the farmers and solve the problem for collection facility location and capacity decisions. We consider 20 different wholesale price options ranging from \$35/dt to \$55/dt. The break-even prices that are closest to these numbers are selected as the associated price options. For example, the analysis for price value \$42/dt is done using the nearest break-even price point to \$42/dt, which is \$42.12/dt.

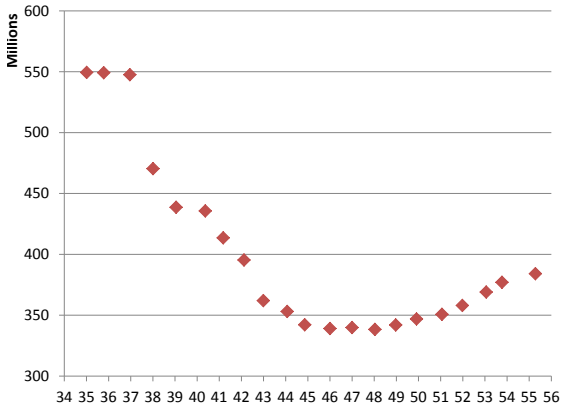


Figure 3.8: Wholesale price (\$/dt) vs. Expected total system cost (\$)

Figure 3.8 shows how the expected total system cost changes according to different wholesale price options. The lowest expected total system cost is achieved when the biomass wholesale price is \$47.74/dt, which we know is the optimal as mentioned earlier. We can see the same result from Figure 3.8 where the expected total system cost is its lowest when price is around \$48/dt. Moreover, we observe that for low price options the expected total system cost is very high due to lack of supply in the system. When low biomass wholesale prices are offered, the farmers do not plant enough switchgrass to satisfy the demand in the system. Therefore, most of

the demand is satisfied via out-sourcing and penalty costs are incurred resulting in a high expected total system cost. As the wholesale price offered to the farmers increases, more expected supply becomes available at farms and utilized to satisfy the demand. Hence, we observe a reduction in the expected total system cost. Nevertheless, after a certain price level, which is \$47.74/dt in this case, offering higher wholesale price values to the farmers becomes less economical. For higher price options, the system has excess supply at farms, which does not contribute to demand satisfaction. Although salvaging the excess supply helps regaining some of the money already spent by the biofuel producer, the expected total system cost starts to increase. As a result, we observe a convex-like expected total system cost function with respect to biomass wholesale price that can be seen from Figure 3.8.

The collection facility location and capacity decisions also vary according to different fixed wholesale prices. For low price values, i.e., \$35/dt-\$42/dt, three or four collection facilities are opened with mostly large capacity levels. On the other hand, the number of collection facilities increases to five or six when biomass wholesale price is set to higher values, i.e., \$48/dt-\$55/dt. For low price values, fewer collection facilities are opened in the system since most of the demand is out-sourced. In this case, the collection facility locations do not affect the transportation costs, since out-sourcing can be performed from any opened collection facility with a fixed unit penalty cost. However, as biomass wholesale price increases, we see more collection facilities with smaller capacity levels to utilize transportation from farms to biorefineries. This is because for high price values, more farmers in different regions produce switchgrass, which turns out to be more economical to transport to biorefineries. As a result, collection facility locations are adjusted according to these new supply configuration.

3.9.2 Different Settings Varying Input Parameters

We change some of the problem parameters and observe how the pricing decision and the supply chain network structure change. We generate four settings by varying input parameter values in the base setting as shown in Table 3.4

Table 3.4: Settings for sensitivity analysis

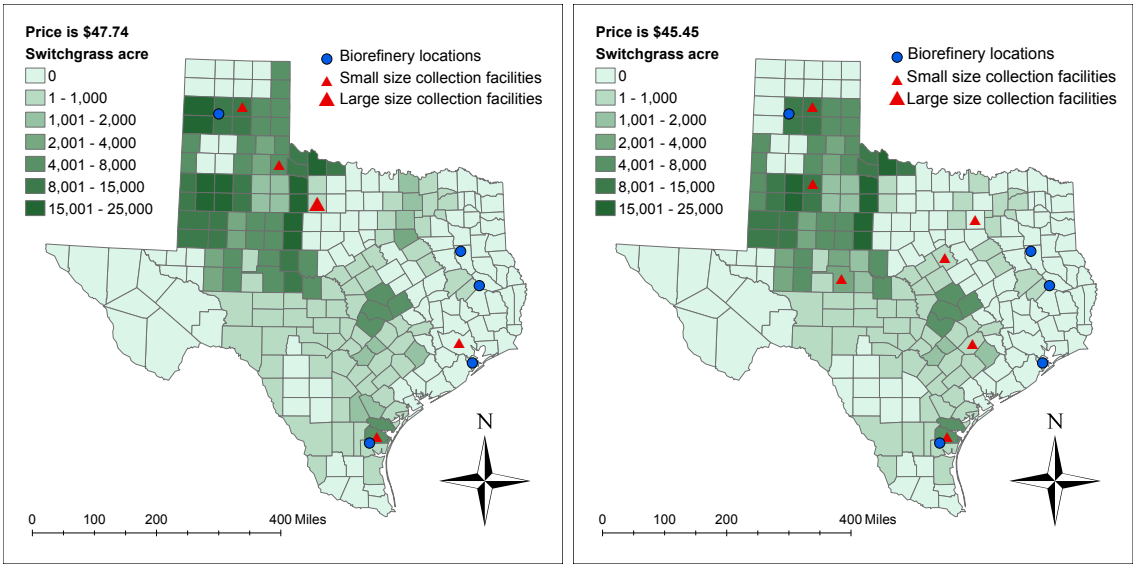
Parameters	S1	S2	S3	S4
Transportation cost ($/dt/mile)	(0.05 - 0.30)	0.15	0.15	0.15
Economies-of-scale discount factor (%)	10	(0 - 50)	10	10
Salvage price ($/dt)	5	5	(0 - 40)	5
Penalty cost ($/dt)	200	200	200	(100 - 300)

In each setting, we alter one input parameter value while keeping the other parameters at their nominal values. In the first setting (S1), we change the unit switchgrass transportation cost. The second setting (S2) alters the economies-of-scale discount factor at collection facilities. Lastly, we change the unit salvage price in setting three (S3) and unit penalty cost in settings four (S4). Figure 3.13 and Figure 3.14 show the effect of these parameters on the price and on the expected total system cost, respectively.

Transportation Cost - S1

In this setting, we analyze the system with different per unit per mile transportation costs ranging from $0.05/dt/mile to $0.30/dt/mile. We observe that for unit transportation costs $0.05/dt/mile, $0.10/dt/mile and $0.15/dt/mile, the optimal price is determined as $47.74/dt, which is the same as in the base setting results. However, for higher unit transportation costs, i.e., $0.20/dt/mile, $0.25/dt/mile

and \$0.30/dt/mile, the optimal wholesale price offered to the farmers decreases to \$45.45/dt. This change in price can be seen in Figure 3.13a. The main reason behind this change is the fact that as unit transportation cost increases, satisfying the demand from distant farms becomes less economical. In other words, it becomes more economical to out-source instead of paying high logistics cost to satisfy the demand. Therefore, the biomass price is lowered and expected biomass supply decreases.



(a) When trans. cost is \$0.05/dt/mile (b) When trans. cost is \$0.30/dt/mile

Figure 3.9: Transportation cost network analysis

Collection facility decisions are also affected when unit transportation cost is changed. Figure 3.9a and Figure 3.9b show the network structures when the unit transportation cost is \$0.05/dt/mile and \$0.30/dt/mile, respectively. For low unit transportation costs, we observe fewer collection facilities in the system. On the other hand, as unit transportation cost increases, more collection facilities with dispersed locations are opened to receive economies-of-scale discount benefits to compensate

the increment in transportation cost. Although less demand is satisfied from farms due to price decrease, more collection facilities are opened to transport the switchgrass supply at farms to biorefineries.

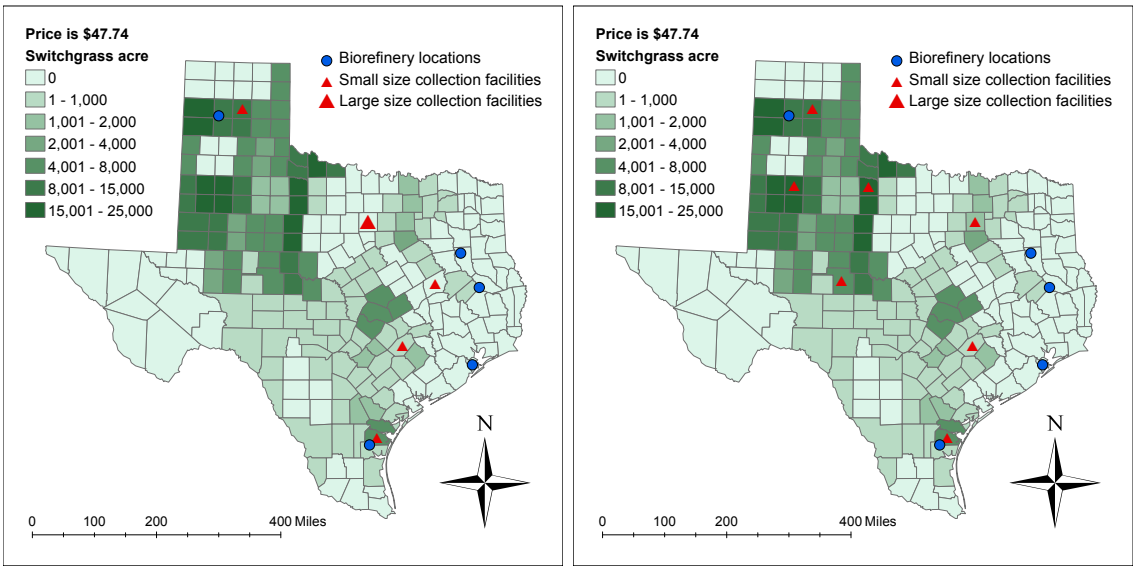
The expected total system cost increases almost linearly as the unit transportation cost increases as shown in Figure 3.14a. When unit transportation cost is doubled from its nominal value, the expected total system cost increases by almost 30%. This shows that logistics cost have a significant effect on the expected total system cost.

Economies-of-scale Discount Factor - S2

In this setting, we change the percentage of the discount that collection facilities provide on the unit transportation cost. We consider different economies-of-scale discount factors ranging from 0% to 50% and observe how the wholesale price and the network structure change.

Our results show that the wholesale price offered to the farmers does not change as the discount factor changes as shown in Figure 3.13b. For all the discount values considered, the price stays the same as in the base setting result, which is \$47.74/dt. This indicates that the discount factor does not have an effect on the biomass wholesale price. However, the network structure changes as the discount factor is altered.

We observe that as the discount factor increases, more collection facilities are opened to enjoy transportation discount benefits. Moreover, we also observe that the collection facility locations tend to get closer to the farms, which plant more switchgrass as the discount factor increases. These results can be seen clearly from Figure 3.10a and Figure 3.10b where optimal collection facility decisions are shown for cases with no discount and 30% discount, respectively. The reason behind this is that as the economies-of-scale discount factor increases, it becomes more econom-



(a) When economies-of-scale discount factor is 0%

(b) When economies-of-scale discount factor is 30%

Figure 3.10: Economies-of-scale discount factor network analysis

ical to open the collection facilities closer to farms with high expected supply since more transportation discount is achieved this way. Notice how similar the network structures are when unit transportation cost is \$0.30/dt and discount factor is 30%.

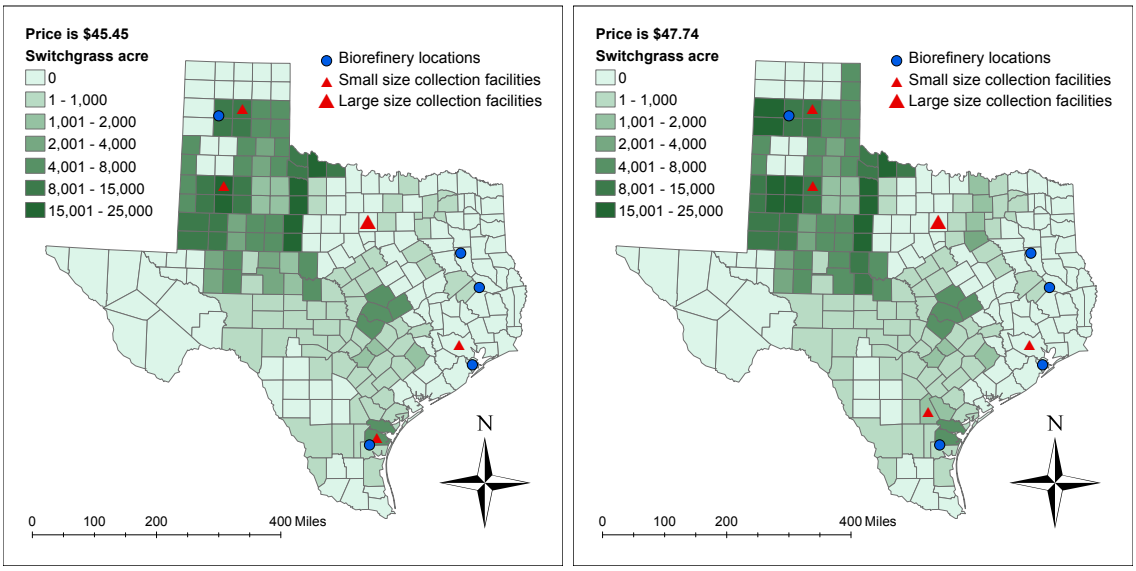
Figure 3.14b shows the relationship between economies-of-scale discount factor and the expected total system cost. Clearly, as the discount factor increases, the expected total system cost decreases due to more transportation cost discount benefits. However, the reduction is not very significant. For example, when the economies-of-scale discount is tripled from its nominal value, only a 4% reduction in the expected total system cost is achieved. Therefore, we can claim that the economies-of-scale discount factor does not have a major effect on the expected total system cost.

Salvage Price - S3

We analyze the effect of unit salvage price in this setting. Different unit salvage prices ranging from \$0/dt to \$40/dt are considered, and their effects on the first-stage decisions and the expected total system cost are observed. Note that when the unit salvage price is \$0/dt, the biofuel producer does not receive any revenue by salvaging excess switchgrass supply.

Figure 3.13c shows how the wholesale price changes for different unit salvage price values. When unit salvage price is \$0/dt, the biomass wholesale price is \$45.45. Notice that this wholesale price is lower than the one obtained in the base setting solution. For unit salvage prices between \$5/dt and \$30/dt, the optimal wholesale price is 47.74/dt. On the other hand, when the unit salvage price is \$35/dt or \$40/dt, the biomass wholesale price increases to \$49.91/dt. Hence, our results indicate that the wholesale price offered to farmers increases as unit salvage price increases. The main reason for this outcome is that the revenue from excess switchgrass supply increases as unit salvage price increases. This gives an incentive to the biofuel producer to offer higher wholesale prices to the farmers since its loss from excess switchgrass is compensated by salvaging.

Figure 3.11a and Figure 3.11b show the optimal collection facility decisions when the unit salvage price is \$0/dt and \$40/dt, respectively. We do not observe major changes in the collection facility decisions as the unit salvage price changes. The network structure we observe in the base setting, where there are two small size collection facilities in North and Southeast Texas and a large size collection facility in between those, remains the same for different unit salvage prices. Therefore, we can say that unit salvage price does not have a major effect on the collection facility decisions.



(a) When salvage price is \$0/dt (b) When salvage price is \$40/dt

Figure 3.11: Salvage price network analysis

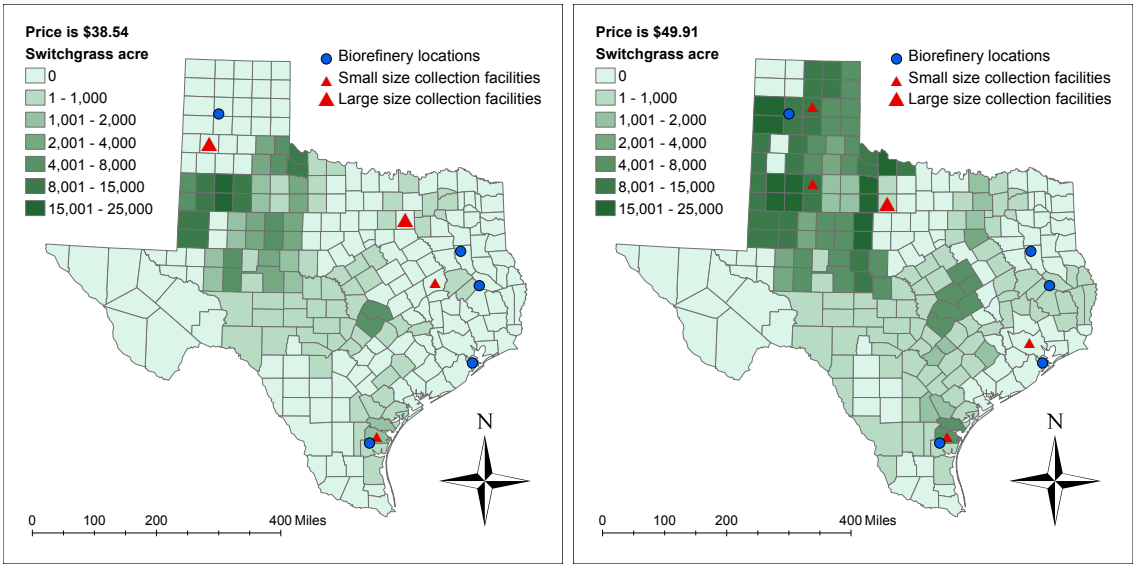
The relationship between unit salvage price and expected total system cost is shown in Figure 3.14c. As the unit salvage price increases, the expected total system cost decreases due to higher revenue coming from salvaging. As in the case of economies-of-scale discount factor, this reduction in expected total system cost is not very significant. To give an example, when the unit salvage price is tripled from its nominal value, the expected total system cost decreases only by 2.5%.

Penalty Cost - S4

In this setting, we consider different unit penalty costs ranging from \$100/dt to \$300/dt. We observe the changes in wholesale price, collection facility locations and capacities and the expected total system cost.

Figure 3.13d shows how the wholesale price changes with respect to different unit penalty costs. As the unit penalty cost increases, the biomass wholesale price increases drastically. When the unit penalty cost is \$100/dt, the wholesale price is

set to $38.54/dt, which is lower than the optimal wholesale price value obtained in the base setting solution. However, when unit penalty cost becomes $300/dt, the optimal wholesale price offered to the farmers is determined as $49.91/dt. This is because for low unit penalty costs, the logistics costs of satisfying one unit demand becomes more than the unit penalty cost. Therefore, instead of spending more money to satisfy the demand via farms, the biofuel producer decides to out-source biomass and bear the penalty cost. On the other hand, for high unit penalty costs, it becomes too costly for the biofuel producer to out-source biomass, hence, higher wholesale price values are offered to farmers to obtain more expected biomass supply. Here, we see a clear trade-off between biomass wholesale price and unit penalty cost.



(a) When penalty cost is $100/dt (b) When penalty cost is $300/dt

Figure 3.12: Penalty cost network analysis

Low wholesale prices result in low expected switchgrass supply and this affects the collection facility decisions. Figure 3.12a shows the optimal collection facility

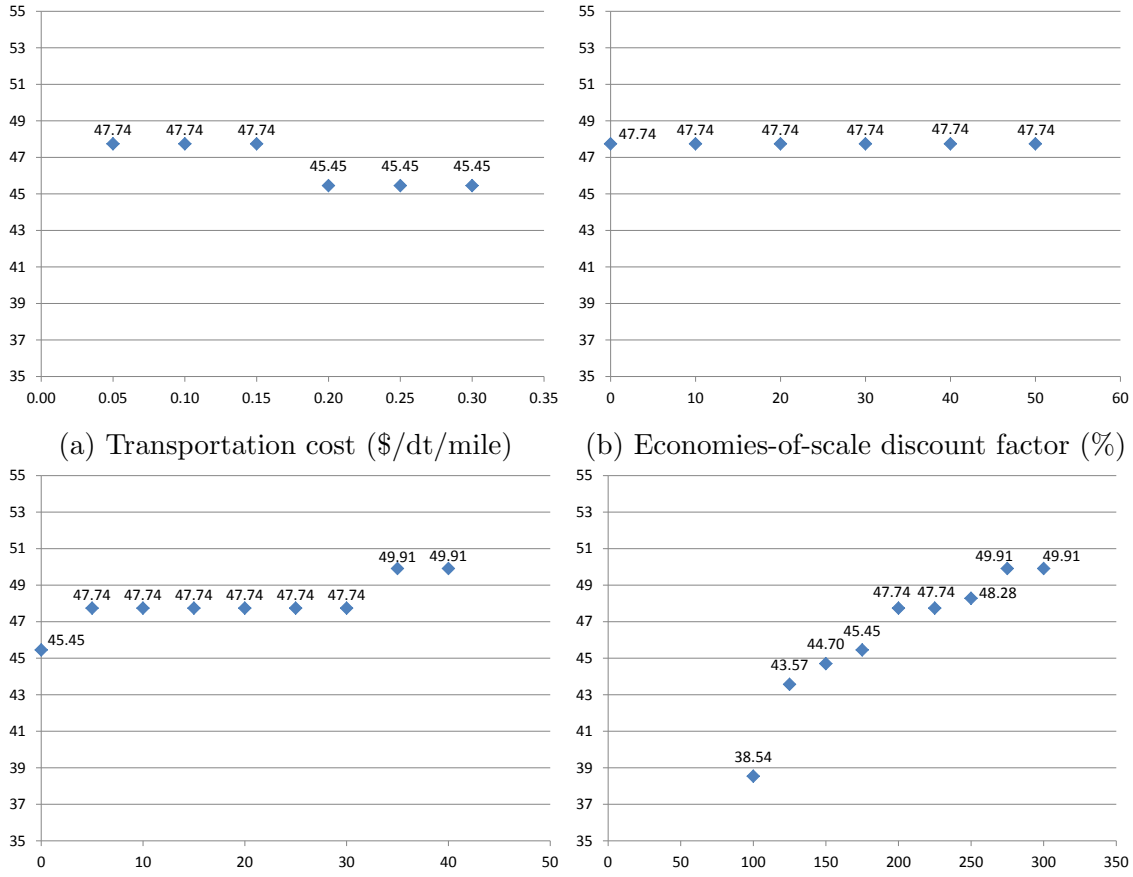
locations and capacity sizes when the unit penalty cost is \$100/dt. Two large and two small size collection facilities are opened that is different than the network structure in the base setting solution. Due to low expected supply, more demand is satisfied via out-sourcing, and hence, fewer collection facilities are required. On the other hand, as the unit penalty cost increases, the network structure looks more like the one obtained in the base setting solution. Figure 3.12b shows the collection facility decisions when the unit penalty cost is \$300/dt and notice that the general network structure is the same as in the base setting solution.

The relationship between unit penalty cost and expected total system cost is shown in Figure 3.14d. Our results indicate that as unit penalty cost increases, expected total system cost also increases in a decreasing fashion. This concave-like shape can be clearly seen in Figure 3.14d. For instance, when unit penalty cost is set to half of its nominal value, the expected total system cost decreases by 17%.

Summary of Results

We summarize our results below:

- The wholesale price offered to farmers is affected significantly when the unit penalty cost changes. Moreover, the unit transportation cost and the unit salvage price have also some effect on the wholesale price but not as much as the unit penalty cost. We do not observe any effect of the economies-of-scale discount factor on the wholesale price.
- The expected total system cost is affected considerably by the unit transportation cost and the unit penalty cost. Although the economies-of-scale discount factor and the unit salvage price affect the expected total system cost, these effects are limited.
- The unit transportation cost, the economies-of-scale discount factor and the



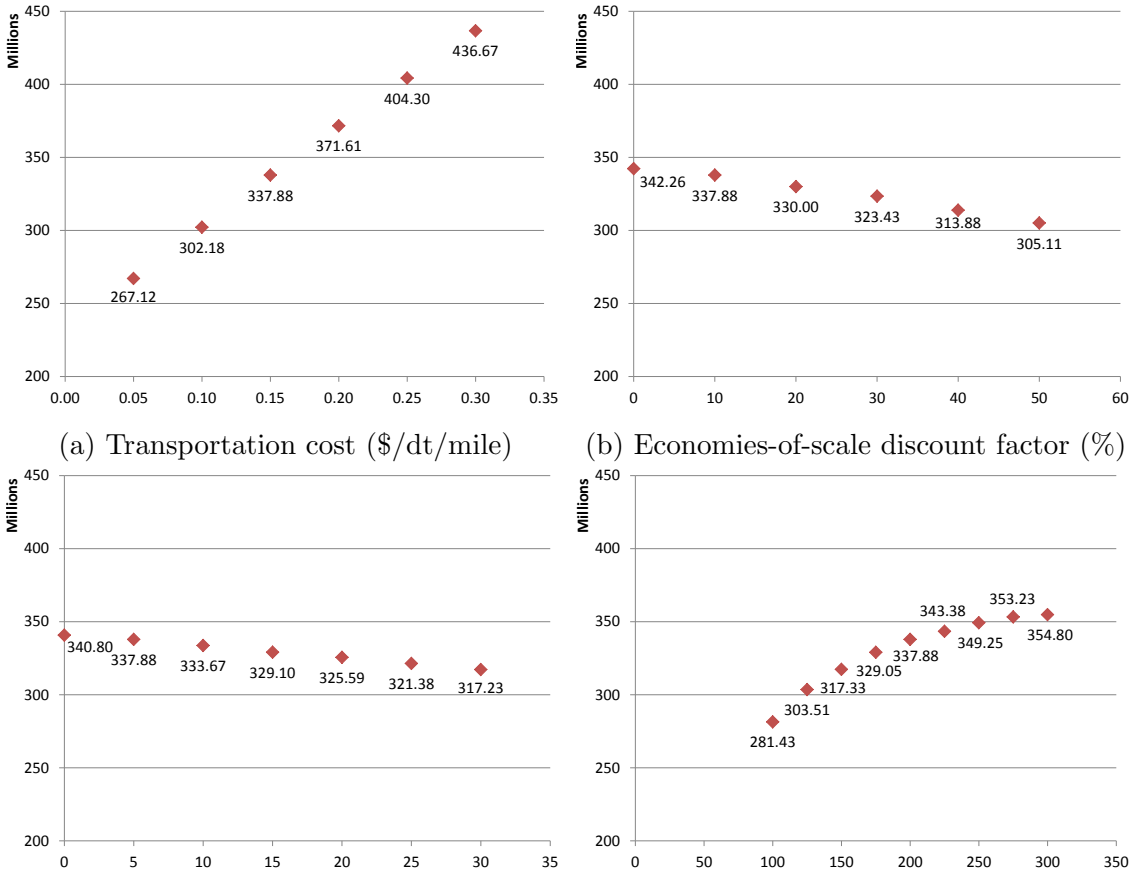
(a) Transportation cost ($/dt/mile)

(b) Economies-of-scale discount factor (%)

(c) Salvage price ($/dt)

(d) Penalty cost ($/dt)

Figure 3.13: Parameter values vs. Wholesale price ($/dt)



(a) Transportation cost ($/dt/mile)

(b) Economies-of-scale discount factor (%)

(c) Salvage price ($/dt)

(d) Penalty cost ($/dt)

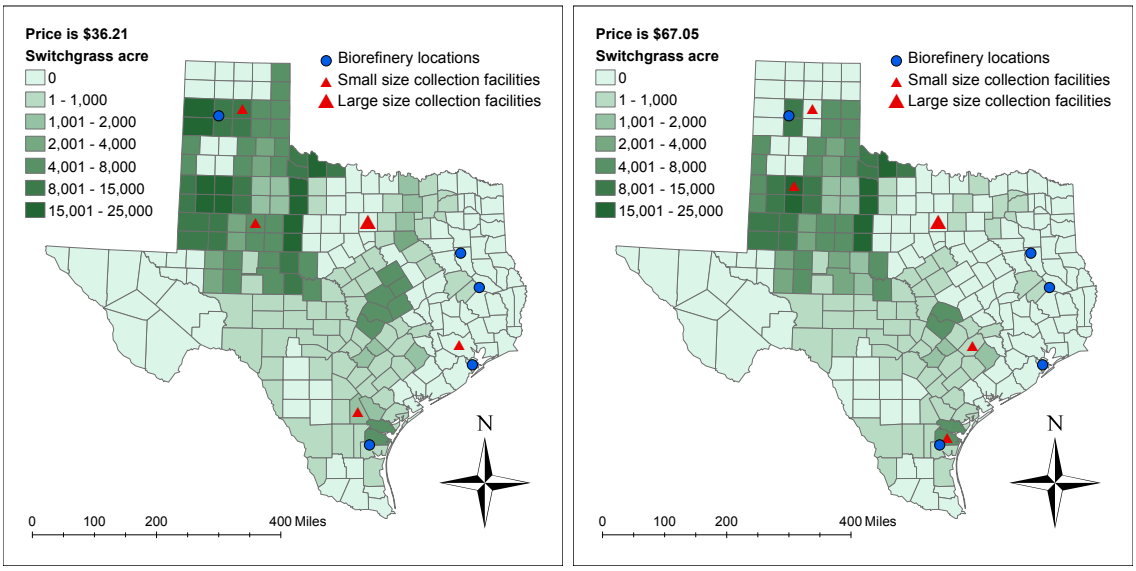
Figure 3.14: Parameter values vs. Expected total system cost ($)

unit penalty cost have an effect on the facility location decisions. As the value of these parameters increase, more collection facilities are opened. When the unit salvage cost is changed on the other hand, the general network structure does not change.

3.9.3 Settings Varying Farmers' Decision Model Parameters

In this setting, we change the farmers' decision model parameter values, specifically the switchgrass yield that farmers estimate (YLD_{i*}) in equation 3.11. As a result, some of the input parameters, i.e., break-even prices for each farmer, change. If farmers estimate a high switchgrass yield that means they are optimistic and willing to take risks to switch their production to switchgrass. On the other hand, if the farmers estimate a low switchgrass yield, meaning they are pessimistic, they become more risk-averse and less willing to switch their production. Note that estimating different values for switchgrass yield does not have an effect on the acre of farmland they switch, i.e., ω values.

For the risk-taker farmers setting, the break-even prices are calculated according to the assumption that each farmer estimates a high switchgrass yield (8dt/acre) at his farm. This will make farmers to switch their production to switchgrass when they are offered lower wholesale prices. Our results show that the optimal wholesale price offered to farmers is \$36.21/dt and the expected total system cost is around \$294 million in this setting. In the risk-averse farmers setting, we calculate the break-even prices based on the assumption that each farmer estimates a low switchgrass yield (4dt/acre) at his farm. With this assumption, the break-even prices increase, meaning that higher wholesale prices should be offered to farmers to convince them to switch their production to switchgrass. Our results show that the optimal biomass wholesale price is \$67.05/dt with an expected total system cost of \$405 million for



(a) Risk-taker farmers setting (b) Risk-averse farmers setting

Figure 3.15: Farmers' decision model network analysis

this setting.

Figure 3.15 shows the network structure of risk-taker and risk-averse farmers settings. Both from Figure 3.15a and Figure 3.15b we see that five collection facilities, four of them are in small size and one in large size, are opened. This is the same number of collection facilities in the base setting result, i.e., risk-neutral farmers, solution. Moreover, the general network structure is also the same where two small size collection facilities are opened in North Texas and also in South and East Texas. A large size collection facility is opened in between these two sets of small size collection facilities. These results show us that although the farmers' decision parameters affect the wholesale price, they do not have a major effect on the network structure.

Figure 3.16 shows the expected total system cost for different wholesale price values under risk-neutral, risk-taker and risk-averse farmers settings. In each setting, the expected total system costs show a similar characteristic by having a convex-like

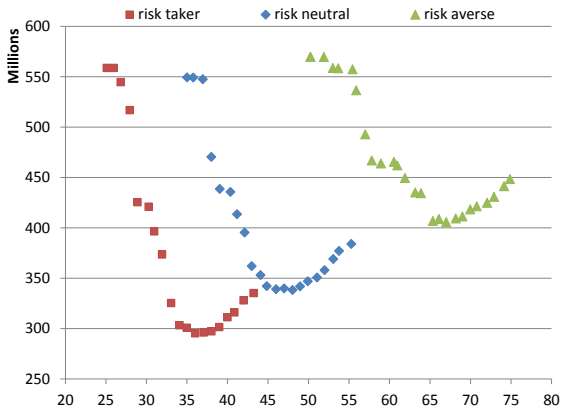


Figure 3.16: Fixed wholesale price values under different farmer decisions ($/dt) vs. Expected total system cost ($)

shape with respect to biomass wholesale price. In all three settings, the expected total system cost is high for low wholesale price values. Of course the measure of "low price" is different in each setting. As the wholesale price increases, first the expected total system cost decreases since farmers plant more switchgrass resulting in an increase in the expected supply. However, as the wholesale price keeps increasing, the biofuel producer starts having more and more excess biomass supply, and hence, the expected total system cost increases.

We observe two major shifts in Figure 3.16, although the general shape of the expected total system cost curves stay the same for different farmer decisions. The first shift is on the x-axis (wholesale price), which increases as farmers get more risk-averse. This is because risk-averse farmers are pessimistic about switchgrass yield and therefore, they need higher wholesale prices to switch their production. The second shift is on the y-axis (expected total system cost), which also increases as farmers become more risk-averse. The main reason for this is the increase in the total expected purchasing cost as price increases.

3.10 Summary & Conclusion

In this section, we present a stochastic program to model farm-to-biorefinery biomass supply chain. Integrated with this supply chain system, we also propose a policy, which is a wholesale price agreement between the parties in the supply chain, i.e., biofuel producer and a group of farmers, to utilize biomass energy crop (switchgrass) supply for biofuel production. To integrate this policy into our problem, we present farmer's decision model, which features the relationship between biomass price and supply. Farmer's decision model enables us to reformulate the overall model into a two-stage linear stochastic program, which simultaneously incorporates the uncertainty in biomass yield and the effects of biomass price on supply, to design the supply chain network. This overall modeling framework allows us to examine the relationship between biomass price and biomass supply chain network structure, which is an area not addressed in the literature.

To solve the reformulated two-stage stochastic program, we utilize a solution methodology based on the L-shaped method, which decomposes the overall problem into smaller subproblems. To improve the algorithmic performance and efficiency, we propose a scenario aggregation approach, which combines scenarios based on their yield values and geographical locations, to generate multiple Benders type cuts to employ in our solution algorithm. Moreover, we further improve the quality of these cuts by solving simple auxiliary problems. Our numerical results show that the proposed multiple-cut approach preforms better than the traditional single-cut approach in terms of computational efficiency. Hence, using our solution methodology, we are able to solve large-scale problem instances in reasonable time. Moreover, we apply an SAA approach to our solution methodology to statistically justify that our solutions are close enough to the true optimal solution of the overall problem, which is

impossible to determine since there are infinitely many scenarios in reality.

Lastly, we conduct a case study in Texas using realistic data to test our model's capabilities and to demonstrate its effectiveness. We devise different settings in which we alter some of the input problem parameter values to deduct insights about the nature of the problem and to investigate the trade-offs between various decisions. Our results indicate that the biomass price is affected drastically by the changes in unit penalty (out-sourcing) cost and the changes in unit transportation cost affect the network structure and the expected total system cost significantly. Moreover, we observe major changes in the expected total system cost and the biomass price when farmers' decision model parameter values are altered.

4. BIO-ENERGY SUPPLY CHAIN NETWORK DESIGN WITH PRICE-BASED SUPPLY AND YIELD UNCERTAINTY

4.1 Introduction

The energy consumption has increased significantly worldwide with the technological and economic developments achieved in recent decades and it will likely to continue increasing in coming years. United States Energy Information Administration [62] projects world energy use to increase 53% by 2035. Almost 30% of the total energy consumption goes into transportation sector and 93% of this is satisfied by fossil fuels [64]. The fact that fossils fuels are not renewable and cause pollution, creates risks and sustainability issues for the future. Although fossil fuels remain the dominant source of energy, renewable energy is projected to be the fastest growing source of primary energy over the next 25 years [62]. Among the renewable energy sources, biofuels have become increasingly popular. Biofuels are sources of clean energy and they are considered as one of future energy sources that can reduce the dependency on fossil fuels. Biofuels cause less carbon emission than fossil fuels and hence, are less harmful to the environment. In fact, research studies show that biofuels decrease vehicle carbon emissions thus making them environmental-friendly [33].

Countries with high energy demands like the U.S. realized this trend in energy consumption and the concerns in environmental sustainability. To respond, in 2005, the U.S. established its first renewable fuel volume mandate, which is called Renewable Fuel Standard (RFS) program and it was created under the Energy Policy Act (EPAct). According to this mandate, 7.5 billion gallons of renewable fuel was required to be blended into gasoline by 2012. This program was extended in 2007

and the amount of renewable fuel to be blended into gasoline was set to 36 billion gallons by 2022 [65].

Currently, majority of the biofuel is produced using edible crops such as corn and sugar cane. Some argue that the production of biofuels using these crops increases food prices and may lead to a global food crises [44]. Therefore, biofuel producers turn their focus on using non-edible sources such as animal and environmental wastes and/or dedicated energy crops, to produce biofuel. Switchgrass, which is an energy crop native to North America, is proved to be a viable source to produce biofuel. A well managed and maintained switchgrass for biofuel production can reach up to 10 feet with a life cycle of 10 to 20 years and an approximate yield around 6 dry tons per acre [22]. However, studies like Jensen et al. [28] and Villamil et al. [71] show that most farmers do not have enough knowledge about switchgrass to adopt it. Therefore, economical incentives should be given to farmers to encourage them to change their production to switchgrass.

For these reasons, a tool is needed to assist biofuel producers to design their biomass-biofuel supply chain network. This tool should simultaneously determine the policies, which give incentives to the farmers to stimulate biomass energy crop production. Moreover, it should also consider the uncertainties in switchgrass supply i.e., yield, caused by weather and environmental conditions, which is an important aspect of biomass-biofuel supply chain design.

In this study, we present a two-stage stochastic program to model farm-to-blending facility biomass-biofuel supply chain system. Our model considers multiple time periods and incorporates the uncertainty in switchgrass yield due to environmental factors. Moreover, we assume there is a wholesale price agreement between the biofuel producer and the regional farmers who are willing to plant switchgrass for biofuel production. According to this agreement, biofuel producer purchases all the

switchgrass produced by farmers with a unit wholesale price determined by itself. The biofuel producer's objective is to maximize its profit considering the revenue coming from biofuel sale to blending facilities and the strategic and logistics costs associated with its biomass-biofuel supply chain. We test our model using realistic data utilized by GIS and implement it on Texas. For different biofuel prices, we examine when the system is profitable or in other words, when it is economically justifiable to invest in biofuel production.

This section is organized as follows: Section 4.2 presents the problem description and the mathematical formulation. In Section 4.3, we present farmers decision model and the reformulated mathematical model. Our solution methodology, which is based on the L-shaped method, is described in Section 4.4. Moreover, the SAA approach employed in the solution procedure is also explained in that section. In Section 4.5, we present the data used in our case study, and our analysis and results. Finally, in Section 4.6, we provide our concluding remarks. We already cover the related literature in biomass-biofuel supply chain context in Section 2.3 and Section 3.2, hence, we do not present a separate literature review section in this section.

4.2 Problem Description & Formulation

We consider a three-level biomass-biofuel supply chain system. These levels are: (i) farms where biomass energy crop, i.e., switchgrass, is produced, (ii) biorefineries where switchgrass is processed and converted to biofuel and (iii) blending facilities, where biofuel, i.e., ethanol, is demanded and mixed with gasoline.

Switchgrass is produced at farms whose locations are known. Switchgrass supply at farms depends on two factors. The first factor is the acreage of land which the farmers plant. The switchgrass acreage at each farm depends on the price which is offered to all farmers by the biofuel producer. By offering this price, the biofuel

producer makes a commitment of buying all the switchgrass supply at these farms for multiple periods of time. Therefore, farmers' planting acreage in these time periods is the same and is determined based on the switchgrass price offered to them at the beginning of the planning horizon. This agrement between the biofuel producer and the farmers is similar to a wholesale price contract. Based on this commitment, all the switchgrass produced at farms is purchased by the biofuel producer, and the excess amount is salvaged. The second factor, which is the switchgrass yield per acre, depends on environmental conditions and is assumed to be unknown. We address this issue by generating random yield scenarios for each farm at each time period in our mathematical formulation.

After harvested at farms, switchgrass is shipped to biorefineries. Biorefineries are biofuel production facilities where switchgrass is processed and their locations and capacities are determined by the decision maker i.e., biofuel producer. All switchgrass shipped to the biorefineries at a time period is processed at the same time period, and converted to biofuel. In other words, we do not consider switchgrass inventory at biorefineries. However, biofuel can be stored at biorefineries to be sold to blending facilities in future time periods. Biorefineries have production and biofuel inventory capacities that limit the biofuel production and inventory amounts for a given time period, respectively.

The blending facility locations and the maximum amount of biofuel they can blend for each time period, i.e., demand upper bound, are known. The biofuel producer sells the produced biofuel at the opened biorefineries to the blending facilities and makes revenue. However, the biofuel producer does not have to fulfill a certain demand level at blending facilities. In other words, the biofuel producer transports biofuel to blending facilities only if it makes profit. If this investment is not profitable for the biofuel producer, biorefineries are not opened and biofuel is not transported

to blending facilities.

We model this problem as a two-stage stochastic program based on scenarios. The biofuel producer makes revenue from selling biofuel to the blending facilities as well as salvaging excess switchgrass. On the other hand, the biofuel producer incurs costs associated with (i) opening biorefineries at the beginning of the planning horizon and (ii) purchasing switchgrass, (iii) transporting switchgrass from farms to biorefineries and biofuel from biorefineries to blending facilities, (iv) processing switchgrass at biorefineries and (v) biofuel inventory in biorefineries, at each time period. The objective of this problem is to maximize the biofuel producer's total profit. In the first-stage, we determine the biorefinery locations and capacities, as well as the switchgrass wholesale price offered to the farmers. These decisions are taken prior to the knowledge of the switchgrass yield for all farms at each time periods. In the second-stage, based on the realized switchgrass supply amounts, switchgrass and biofuel transportation, switchgrass salvaging and biofuel inventory decisions are taken for all time periods.

Based on the problem description presented above, the two-stage stochastic program is presented below in its Deterministic Equivalent Program (DEP) format. The DEP is the large-scale MIP version of the SP where each constraint associated with second-stage decisions is written explicitly for each scenario. Moreover, each second-stage decision variable is also explicitly defined for each scenario in this formulation.

Before, presenting the mathematical model, we first introduce the notation used:

Sets:

$\mathcal{I}$	set of farms indexed by i
$\mathcal{K}$	set of biorefineries indexed by k
$\mathcal{Z}$	set of blending facilities indexed by z
$\mathcal{L}$	set of capacity (size) levels indexed by l
$\mathcal{T}$	set of time periods indexed by t
$\mathcal{S}$	set of yield scenarios indexed by s

Parameters:

π_t^b	unit biofuel price at time period t
$\Gamma_i(\pi^m)$	function that gives the acre of switchgrass planted in farm i according to price π^m
Φ_{it}^s	switchgrass yield per acre in farm i at time period t according to scenario s
f_{kl}	fixed cost of opening and operating a biorefinery of size l at location k
k_l	biofuel production capacity of size l biorefinery
e_l	biofuel inventory capacity of size l biorefinery
u	unit switchgrass salvaging price
c	unit switchgrass processing cost
h	unit biofuel holding cost for a time period
d_{zt}^U	upper bound quantity of biofuel demanded at blending facility z
t_{ik}^m	per unit switchgrass transportation cost from farm i to biorefinery k
t_{kz}^b	per unit biofuel transportation cost from biorefinery k to blending facility z
p^s	probability of scenario s
v	unit switchgrass conversion rate to biofuel

Decision Variables:

Z_{kl} takes value 1, if a biorefinery of size l is opened at location k and takes value 0, otherwise

π^m unit switchgrass wholesale price

X^s_{ikt} quantity of switchgrass shipped from farm i to biorefinery k at time period t for scenario s

Y^s_{kzt} quantity of biofuel shipped from biorefinery k to blending facility z at time period t for scenario s

I^s_{kt} quantity of biofuel kept in inventory in biorefinery k at time period t for scenario s

W^s_{it} quantity of switchgrass salvaged at farm i at time period t for scenario s.

According to the notation introduced, the two-stage stochastic program is formulated in its DEP form as follows:

$$\begin{aligned}\text{Max} \quad & -\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} f_{kl}\, Z_{kl} + \sum_{s\in\mathcal{S}} p^s \Bigg[\sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} \pi^b_t\, Y^s_{kzt} + \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} u\, W^s_{it} \\ & - \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} \Phi^s_{it}\, \pi^m\, \Gamma_i(\pi^m) - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} c\, X^s_{ikt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} t^m_{ik} X^s_{ikt} \\ & - \sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} t^b_{kz} Y^s_{kzt} - \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} h\, I^s_{kt} \Bigg] \end{aligned} \tag{4.1}$$

subject to

$$\sum_{l\in\mathcal{L}} Z_{kl} \leq 1 \qquad \forall\, k \in \mathcal{K} \tag{4.2}$$

$$v \sum_{i\in\mathcal{I}} X^s_{ikt} \leq \sum_{l\in\mathcal{L}} k_l\, Z_{kl} \qquad \forall\, k \in \mathcal{K}, \forall\, t \in \mathcal{T}, \forall\, s \in \mathcal{S} \tag{4.3}$$

$$I^s_{kt} \leq \sum_{l\in\mathcal{L}} e_l\, Z_{kl} \qquad \forall\, k \in \mathcal{K}, \forall\, t \in \mathcal{T}, \forall\, s \in \mathcal{S} \tag{4.4}$$

$$\sum_{k\in\mathcal{K}} Y^s_{kzt} \leq d^U_{zt} \qquad \forall\, z \in \mathcal{Z}, \forall\, t \in \mathcal{T}, \forall\, s \in \mathcal{S} \tag{4.5}$$

$$\sum_{k\in\mathcal{K}} X^s_{ikt} + W^s_{it} = \Gamma_i(\pi^m)\, \Phi^s_{it} \qquad \forall\, i \in \mathcal{I}, \forall\, t \in \mathcal{T}, \forall\, s \in \mathcal{S} \tag{4.6}$$

$$v \sum_{i\in\mathcal{I}} X^s_{ikt} + I^s_{k(t-1)} = \sum_{z\in\mathcal{Z}} Y^s_{kzt} + I^s_{kt} \qquad \forall\, k \in \mathcal{K}, \forall\, t \in \mathcal{T}, \forall\, s \in \mathcal{S} \tag{4.7}$$

$$u \leq \pi^m \tag{4.8}$$

$$X^s_{ikt},\, Y^s_{kzt},\, W^s_{it},\, I^s_{kt} \geq 0 \tag{4.9}$$

$$Z_{kl} \in \{0, 1\} \qquad \forall\, k \in \mathcal{K}, \forall\, l \in \mathcal{L}. \tag{4.10}$$

The objective function (4.1) maximizes the expected profit of the biofuel producer over the planning horizon. The first term in (4.1) is the fix cost associated with biorefineries. The second term includes the revenues and costs associated with the decisions taken in each time period over the planning horizon. Notice that each term within the parenthesis is summed over all the time periods. First two terms within the parenthesis correspond to the revenue coming from selling biofuel and salvaging switchgrass, respectively. Biomass purchasing cost is stated in the third term. The fourth term is for switchgrass processing cost, also referred as the production cost.

Switchgrass and biofuel transportation costs are addressed in the fifth and sixth terms, respectively. The last term in (4.1) corresponds to biofuel inventory cost.

Constraint (4.2) makes sure that only one biorefinery is opened at a candidate location. Constraint (4.3) is to limit the biofuel production at biorefineries according to their associated production capacities. Biofuel inventory capacity for each time period is stated in constraint (4.4). The biofuel demand upper bound constraint at each blending facility is in (4.5). Constraint (4.6) is the supply constraint. Biofuel inventory balance constraint is stated (4.7). The switchgrass wholesale price offered to the farmers must be greater than the unit salvage price and this is covered in constraint (4.8). Lastly, continuous and binary decision variables are defined in (4.9) and (4.10), respectively.

In this DEP formulation of our SP, all the second-stage decision variables and their associated constraints are explicitly written for each scenario. Therefore, the problem size increases drastically as the scenario set $\mathcal{S}$ increases. This makes it very challenging to solve this problem for large number of scenarios. Hence, a solution algorithm is essential to solve this problem. We will present our proposed solution algorithm in Section 4.4.

Moreover, the nature of the mathematical formulation presented in (4.1)-(4.10) depends on the structure of the Γ_i function. If this function is non-linear for example, the objective function (4.1) and the right hand side of constraint (4.6) become non-linear. As a result, the problem becomes very difficult to solve using traditional methods. On the other hand, if Γ_i is linear, whole problem becomes linear and can be solved rather easily. We will address the structure of this function in Section 4.3 and explain how we construct Γ_i for each farmer i.

4.3 Farmers' Decision & Problem Reformulation

The Γ function has a important affect on the problem structure. Therefore, we first address how this function is obtained for each farmer by constructing farmer's decision model.

4.3.1 Farmers' Decision

To model the relationship between switchgrass supply and price, we employ the approach used in Downing and Graham [17]. In this approach, we assume that a farmer makes his switchgrass planting decision based solely on his expected profit. In other words, the farmer adopts switchgrass if the expected profit of switchgrass is higher than at least one of the crops that the farmer currently plants. This is the same assumption that we used in Section 3.4. However in this case, once the planting decision is determined, we assume that the farmer uses this decision for all the time periods in the planning horizon. In other words, the acreage on which the farmer plants switchgrass is determined at the beginning of the planning horizon and stays the same for all time periods.

Let's use the same notation and parameters introduced in Section 3.4. We know that for each farmer i and every crop r equation (3.11) must be true. In essence, this equation states that the switchgrass break-even price for a crop is the price value which provides the same profit as that crop provides. For each crop the farmers currently plant, equation (3.11) is solved and break-even prices are obtained.

As illustrated in Figure 3.3, the Γ function has a stepwise structure. Using the nature of this function, we can state that the optimal wholesale price offered to a farmer should be one of his break-even prices as shown in Proposition 3.4.1. This is because offering prices between the break-even price values does not have an affect on the acreage that farmers plant but increase the purchasing cost for the

biofuel producer. This fact can be generalized to a set of farmers as it is done in Corollary 3.4.2. Thus, we can state that the optimal price offered to a group of farmers is one of the break-even prices of a farmer in the group. In fact, this reduces the number of candidate prices to a finite set and we will use this to reformulate our problem in Section 4.3.2

4.3.2 Problem Reformulation

We know that the optimal wholesale price offered to farmers must be one of their break-even prices. Therefore, we have a finite number of candidate prices and each can be represented by a binary variable. We follow a similar notation used in Section 3.5 to reformulate our model. Let $\mathcal{F}$ be the set of all possible unit switchgrass prices i.e., possible break-even prices. Moreover, let P_f be a binary decision variable which takes value 1 if price $f \in \mathcal{F}$ is selected, and takes value 0 otherwise. Furthermore, π_f^m is the price value of f. We also present the parameter ω_{if}, which is the acreage planted by farmer i when price f is offered to him, i.e., $\Gamma_i(\pi_f^m)$.

In addition, we know that biofuel producer purchases all the switchgrass produced by the farmers. Therefore, the expected switchgrass purchasing cost can be taken out of the parenthesis and placed as a separate term in the objective function (4.1). Using these modifications, the reformulated DEP is presented below:

$$\begin{aligned}\text{Max} \quad & -\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} f_{kl}\, Z_{kl} - \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}}\sum_{t\in\mathcal{T}} p^s \Phi^s_{it}\omega_{if}\,\pi_f^m P_f + \sum_{s\in\mathcal{S}} p^s \Bigg[\sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} \pi_t^b\, Y^s_{kzt} \\ & + \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} u\, W^s_{it} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} c\, X^s_{ikt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} t^m_{ik} X^s_{ikt} - \sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} t^b_{kz} Y^s_{kzt} \\ & - \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} h\, I^s_{kt}\Bigg] \end{aligned} \tag{4.11}$$

subject to $(4.2), (4.3), (4.4), (4.5), (4.7), (4.9), (4.10)$,

$$\sum_{f \in \mathcal{F}} P_f = 1 \tag{4.12}$$

$$\sum_{k \in \mathcal{K}} X^s_{ikt} + W^s_{it} = \sum_{f \in \mathcal{F}} \Phi^s_{it} \omega_{if} P_f \qquad \forall\, i \in \mathcal{I}, \forall\, t \in \mathcal{T}, \forall\, s \in \mathcal{S} \tag{4.13}$$

$$P_f \in \{0, 1\} \qquad \forall\, f \in \mathcal{F} \tag{4.14}$$

Note that the second term of objective function (4.11) is the expected switchgrass purchasing cost, which is written outside of the parenthesis compared to the earlier formulation. Constraint (4.12) makes sure only one of the break-even price values is selected, and the corresponding binary decision variable is defined in (4.14). Constraint (4.13) is a different representation of constraint (4.6) using the new binary decision variable P_f and parameter ω_{if}. We no longer require constraint (4.8), since set $\mathcal{F}$ includes only break-even price values that satisfy this constraint. Hence, constraint (4.8) is taken out of the formulation.

4.4 Solution Methodology

In this section, we explain our methodology to solve the large-scale MIP reformulation. First, we present our solution algorithm based on L-shaped method, which decomposes the overall problem. Secondly, we introduce the SAA method, which solves the same problem with different set of scenarios to statistically justify the goodness of the solution.

4.4.1 L-shaped Method

As explained in Section 3.6, the L-shaped method, in essence, is the same as the Benders Decomposition algorithm. In the L-shaped framework, the DEP model introduced in Section 4.3 is decomposed into a master and a subproblem. The

decision variables in the master problem correspond to the first-stage decisions in our original two-stage stochastic program. On the other hand, the subproblem addresses the decisions taken in the second-stage for each time period. Since the subproblem is written for each scenario explicitly, it can be further decomposed into smaller subproblems where each subproblem corresponds to a single scenario. Therefore, in each iteration of the L-shaped method, the subproblem is solved for each scenario separately. Using the dual solutions of these subproblems, a Benders type cut is generated and added to the master problem.

The L-shaped Subproblem (LsSP)

We obtain the subproblem by assuming given biorefinery ($\hat{Z}_{kl}$) and price ($\hat{P}_f$) decisions in our DEP. The overall LsSP is presented below:

$$\begin{aligned}\text{Max} \quad & \sum_{s\in\mathcal{S}} p^s \Bigg[\sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} \pi^b_t\, Y^s_{kzt} + \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} u\, W^s_{it} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} c\, X^s_{ikt} \\ & - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} t^m_{ik} X^s_{ikt} - \sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} t^b_{kz} Y^s_{kzt} - \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} h\, I^s_{kt} \Bigg] \end{aligned} \tag{4.15}$$

subject to (4.5), (4.7), (4.9),

$$v \sum_{i\in\mathcal{I}} X^s_{ikt} \le \sum_{l\in\mathcal{L}} k_l\, \hat{Z}_{kl} \qquad \forall\, k\in\mathcal{K}, \forall\, t\in\mathcal{T}, \forall\, s\in\mathcal{S} \tag{4.16}$$

$$I^s_{kt} \le \sum_{l\in\mathcal{L}} e_l\, \hat{Z}_{kl} \qquad \forall\, k\in\mathcal{K}, \forall\, t\in\mathcal{T}, \forall\, s\in\mathcal{S} \tag{4.17}$$

$$\sum_{k\in\mathcal{K}} X^s_{ikt} + W^s_{it} = \sum_{f\in\mathcal{F}} \Phi^s_{it}\omega_{if}\hat{P}_f \qquad \forall\, i\in\mathcal{I}, \forall\, t\in\mathcal{T}, \forall\, s\in\mathcal{S} \tag{4.18}$$

The overall LsSP can be decomposed and written for each scenario separately. Without loss of generality, we define dual variables λ^s_{zt}, θ^s_{kt}, α^s_{kt}, β^s_{kt} and μ^s_{it} corresponding to scenario s and associated with constraints (4.5), (4.7), (4.16), (4.17)

and (4.18), respectively. Hence, for scenario s, the dual subproblem LsDSP has the following structure:

$$\begin{aligned}\text{Min} \quad & \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} k_l\,\hat{Z}_{kl}\alpha^s_{kt} + \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} e_l\,\hat{Z}_{kl}\beta^s_{kt} \\ & + \sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} d^U_{zt}\lambda^s_{zt} + \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}}\sum_{f\in\mathcal{F}} \Phi^s_{it}\omega_{if}\hat{P}_f\mu^s_{it} \qquad (4.19)\end{aligned}$$

subject to

$$\lambda^s_{zt} - \theta^s_{kt} \geq \pi^b_t - t^b_{kz} \qquad \forall\, k\in\mathcal{K}, \forall z\in\mathcal{Z}, \forall\, t\in\mathcal{T} \qquad (4.20)$$

$$v\,\alpha^s_{kt} + \mu^s_{it} + v\,\theta^s_{kt} \geq -c - t^m_{ik} \qquad \forall\, i\in\mathcal{I}, \forall\, k\in\mathcal{K}, \forall\, t\in\mathcal{T} \qquad (4.21)$$

$$\mu^s_{it} \geq u \qquad \forall\, i\in\mathcal{I}, \forall\, t\in\mathcal{T} \qquad (4.22)$$

$$\beta^s_{kt} + \theta^s_{k(t+1)} - \theta^s_{kt} \geq -h \qquad \forall\, k\in\mathcal{K}, \forall\, t\in\mathcal{T} \qquad (4.23)$$

$$\alpha^s_{kt},\ \beta^s_{kt},\ \lambda^s_{zt} \geq 0 \qquad (4.24)$$

$$\mu^s_{it}, \theta^s_{kt} \quad \text{unrestricted.} \qquad (4.25)$$

Let $\mathcal{E}^s$ represent the set of all extreme points in the LsDSP polyhedron given by (4.20)-(4.25) for scenario s. Moreover, let $\lambda^{e^s}_{zt}$, $\theta^{e^s}_{kt}$, $\alpha^{e^s}_{kt}$, $\beta^{e^s}_{kt}$, $\mu^{e^s}_{it}$ and η^{e^s} denote the dual variables and the objective value correspond to the extreme point $e^s \in \mathcal{E}^s$, respectively. In addition, we let η^{*^s} and η^* be the optimal objective values for the LsSP corresponding to scenario s and the overall LsSP, respectively. We know that $\eta^{e^s} \geq \eta^{*^s}$, $\forall e^s \in \mathcal{E}^s$ must be true for each scenario s, since e^s is a feasible point, i.e., an upper bound solution of LsSP for scenario s. Therefore, we can reformulate

LsDSP as $\max\{\eta^s : \eta^{e^s} \geq \eta^s,\ \forall e^s \in \mathcal{E}^s\}$ where

$$\eta^{e^s} = \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} k_l\, \hat{Z}_{kl}\alpha^{e^s}_{kt} + \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}} e_l\, \hat{Z}_{kl}\beta^{e^s}_{kt} + \sum_{z\in\mathcal{Z}}\sum_{t\in\mathcal{T}} d^U_{zt}\lambda^{e^s}_{zt} + \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}}\sum_{f\in\mathcal{F}} \Phi^s_{it}\omega_{if}\hat{P}_f\mu^{e^s}_{it}. \tag{4.26}$$

Expression (4.26) is only for a single scenario. If we consider the expectation of all scenarios, the overall LsDSP can be restated as $\max\{\eta : \eta^e \geq \eta,\ \forall e \in \mathcal{E}\}$ where $\eta^e = \sum_{s\in\mathcal{S}} p^s\eta^{e^s}$.

Reformulation of the L-shaped Master Problem (LsMP)

Using the representation presented above, we write the reformulation of the original problem as follows:

$$\text{Max} \quad -\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}} f_{kl}\, Z_{kl} - \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}}\sum_{t\in\mathcal{T}} p^s\Phi^s_{it}\omega_{if}\,\pi^m_f P_f + \eta \tag{4.27}$$

$$\text{subject to } (4.2), (4.10), (4.12), (4.14)$$

$$\eta \leq \eta^e \qquad \forall\, e \in \mathcal{E} \tag{4.28}$$

$$\eta \quad \text{unrestricted.} \tag{4.29}$$

Constraint (4.28) is written for all extreme points of the LsDSP polyhedron, i.e., set $\mathcal{E}$, which is not available beforehand. Therefore, the reformulation considers only a subset of the constraints in (4.28) and thus, is a relaxed problem of the overall problem. This means that the solution to this relaxed problem provides an upper bound to the overall problem. These constraints are constructed and added in a delayed fashion, one at a time in each iteration of the L-shaped method. The

proposed solution methodology is outlined in Algorithm 3.

Algorithm 3 L-shaped Method

initialize $\epsilon = 0.02$, $\texttt{optgap} = 1.0$, $\texttt{LB} = -\infty$, $\texttt{UB} = \infty$, $\texttt{LB}^{\texttt{best}} = -\infty$
while $(\texttt{optgap} > \epsilon)$ **do**
  Solve **LsMP** and obtain the optimum objective value Z_{LsMP} and its solution vectors $\hat{\mathbf{Z}}$ and $\hat{\mathbf{P}}$ and the auxiliary variable value $\hat{\eta}$
  Set $\texttt{UB} = Z_{LsMP}$
  for All scenarios $s \in \mathcal{S}$ **do**
    Substitute $\hat{Z}_{kl}$ and $\hat{P}_f$ and solve **LsSP**
    Store the optimal objective value Z^s_{LsSP} and associated dual variable solutions
  end for
  Generate a single Benders type cut as in (4.28) and add it to **LsMP**
  Set $\texttt{LB} = Z_{LsMP} - \hat{\eta} + \sum_{s\in\mathcal{S}} p^s Z^s_{LsSP}$
  if $\texttt{LB} > \texttt{LB}^{\texttt{best}}$ **then**
    $\texttt{LB}^{\texttt{best}} = \texttt{LB}$
  end if
  $\texttt{optgap} = (\texttt{UB} - \texttt{LB}^{\texttt{best}})/\texttt{LB}^{\texttt{best}}$
end while
return $\texttt{LB}^{\texttt{best}}$ and its corresponding solution vectors $\hat{\mathbf{Y}}$ and $\hat{\mathbf{P}}$

4.4.2 SAA Approach

Since it is impossible to solve our stochastic program with all possible scenarios, we implement the Sample Average Approximation (SAA) approach, which uses a subset of scenarios to solve the original problem as previously mentioned in Section 3.8.2. We implement the same approach that is utilized in the previous section. However, notice that in this section we have a maximization problem that is different than the model presented in Section 3, which is a minimization problem. Therefore, SAA lower and upper bound problems in the previous implementation correspond to SAA upper and lower bound problems, respectively in this case.

As it is also discussed in Section 3.8.2, two separate problems are solved within the SAA framework: (i) SAA upper bound problem (SAAup) and (ii) SAA lower bound problem (SAAlw). SAAup solves the original problem with a small set of scenarios, and provides an upper bound on the overall problem. In the SAA framework, SAAup is solved several time, i.e., m times, and the mean upper bound value (μ) is obtained. Using the Central Limit Theorem (CLT), a confidence interval (CI) can be constructed around this mean with a certain percentage, e.g., 95% is used in our calculations. This means that the true upper bound to the overall problem lies within that confidence interval with %95 probability.

SAAlw solves the problem with a large set of scenarios for fixed first-stage decisions. In SAA framework, the first-stage decisions corresponding to the solution with highest objective value in the SAAup are fixed. The problem is then solved with a large set of scenarios and a lower bound on the overall problem is obtained. Similar to SAAup, SAAlw is also solved m times, and the mean lower bound value (μ) is obtained and a confidence interval is constructed.

After solving SAAup and SAAlw, and determining the corresponding confidence intervals, the confidence interval gap is calculated. This gap represents the difference between the maximum and the minimum values of the upper bound and lower bound confidence intervals. As the confidence interval gap gets smaller, the upper and lower bound estimates get closer to each other. In other words, the lower bound solution we obtain gets closer to the true optimal value of the overall problem.

We experimented the SAA approach with different scenario numbers (200, 300, 400 and 500) and batch sizes (m=15 and m=25). Table 4.1 shows the mean values and standard deviations for SAAup and SAAlw. We also provide the confidence interval gap, which shows the closeness of the upper and lower bound confidence intervals.

Table 4.1: SAA upper bound problem and SAA lower bound problem results for different batch sizes and scenario numbers

Batch size (m)		Scenario #		Ave (μ)		Std (σ)		Gap (%)
SAAlw	SAAup	SAAlw	SAAup	SAAlw	SAAup	SAAlw	SAAup	CI
15	15	4000	200	138,579,533	138,267,533	1,291,488	4,612,515	3.43
15	15	6000	300	134,905,667	137,238,000	407,109	3,953,705	3.37
15	15	8000	400	136,420,467	137,421,200	587,720	3,661,229	2.73
15	15	10000	500	138,360,533	137,349,933	546,997	2,237,324	1.78
25	25	4000	200	136,597,480	138,161,560	1,525,319	4,760,800	2.96
25	25	6000	300	138,585,880	137,207,960	1,146,315	4,202,948	2.56
25	25	8000	400	138,593,880	139,047,000	703,618	3,799,397	2.17
25	25	10000	500	138,598,840	138,269,960	916,528	2,646,295	1.51

Table 4.1 shows that as the number of scenarios or the batch sizes increase, the confidence interval gap decreases. We observe that only when 500 scenarios for SAAup is used, the confidence interval gap is lower than 2%. For all the other scenario numbers, CI gaps are larger than 2%. Moreover, increasing the batch sizes decreases the confidence interval gap. However, due to time limitations, we decide to solve 15 instances in both SAAup and SAAlw instead of solving more instances. Therefore, we select 500 scenarios for SAAup and 10,000 scenarios for SAAlw and solve 15 instances (batch sizes) in each problem to determine our solution. This configuration provides us with solutions that are at most 1.78% apart with 95% statistical confidence. According to our results, the objective function is around \$138 million, which means the biofuel producer makes that much profit over the 5 year planning horizon.

4.5 Case Study

To demonstrate the capabilities and the applicability of our model, we conduct a case study in Texas using realistic data from the literature. All the instances are

solved implementing the L-shaped based solution algorithm presented in Section 4.4 until 2% optimality gap is reached. Moreover, we utilize an SAA approach solving 15 instance batches with 500 and 10,000 scenarios for SAAup and SAAlw, respectively.

4.5.1 Data Gathering

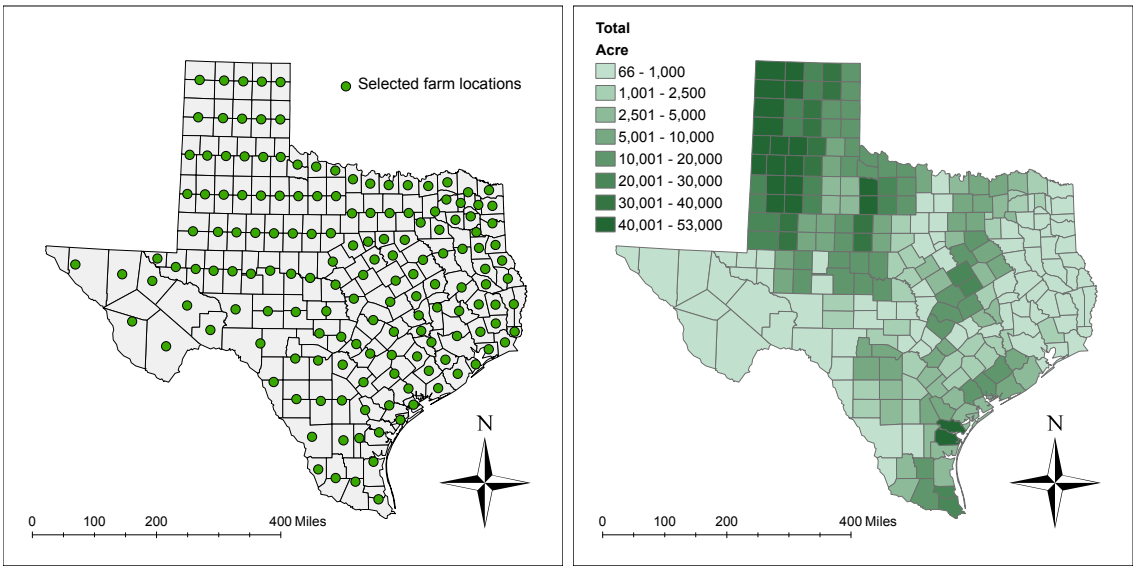
In this section, we explain how we obtained the data used in the case study. We test our model on a realistic data generated in the state of Texas utilizing GIS and considering 5 year planning horizon.

Switchgrass Supply & Farms

To determine the supply points, we first locate the geographical center point of each county in Texas using GIS. Then, some of these counties are paired together by finding the mid-point between them to create a smaller and more representative set of supply points. Each of these supply points corresponds to a farm and as a result, we determine 155 farms in Texas. Moreover, we divide these farms into three groups according to their geographical locations, i.e., Northwest farms, Northeast farms and Southeast farms as we present in Section 3.7. Figure 4.1a shows the selected 155 farms and recall that Figure 3.4b shows the associated farm groups.

We use switchgrass as the biomass energy crop in this case study. Switchgrass supply amounts at each farm are determined according to the farmer's decision model presented in Section 4.3. Same four major crops in Texas, i.e., corn, sorghum, cotton and wheat are selected for this case study. Figure 4.1b shows the total acre currently planted for these crops. In other words, Figure 4.1b presents the total available acre that can be switched to switchgrass production.

We employ the same data explained in Section 3.7.1 from Texas A&M Agrilife Extension Service [52] and United States Department of Agriculture [54, 58], to determine the necessary parameters for equation (3.11). Moreover, as in the previous



(a) Selected farm locations (b) Total available acre

Figure 4.1: Selected farm locations and total available acre

study, we use \$5/dt and \$20/dt as unit salvage price and unit processing cost, respectively. Furthermore, unit switchgrass transport cost is taken as \$0.15/mile/dt and conversion rate for switchgrass is assumed to be 75 gallons/dt.

Yield Scenario Generation

We assume that there are two switchgrass yield levels, high (H) and low (L) for each farm group. For each farm that is in a low yield level group, we use a random variable U[2,6] to generate its switchgrass yield. Similarly, for farms that are in high yield level groups, a random variable U[6,10] is generated. Therefore, for a single time period, we have 8 different yield combinations, i.e., 2^3, similar to the "scenario types" presented in Table 3.1. To give an example, the yield combination (H,L,H) represents high yields for the farms in groups one and three, and low yields for farms in group two.

To construct each scenario we generate a random integer variable between $[1, 8]$

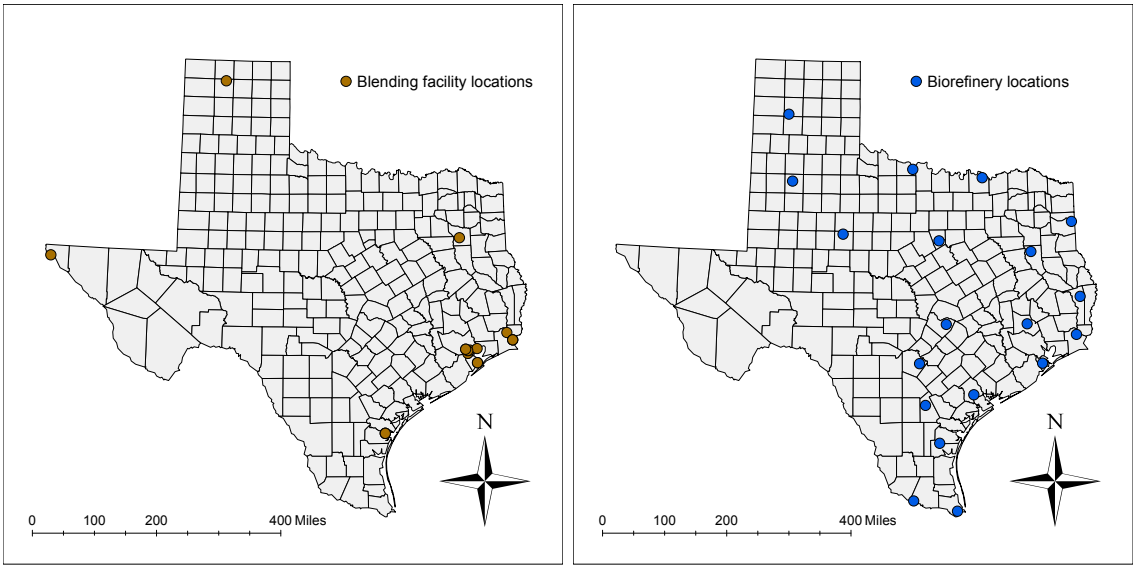
for each time period t. This set of random variables, each integer variable corresponding to a yield combination for a time period, represent a "scenario type". Notice that "scenario type" definition in this section is different than the one presented in Section 3.7.2. For this problem a "scenario type" includes yields for every time period, whereas for the problem presented in Section 3, this definition corresponds only to a single period. For example, the random variable set (1,2,3,4,5) represents a scenario in which we have (H,H,H), (H,H,L), (H,L,H), (H,L,L) and (L,H,H) for time periods one, two, three, four and five, respectively. Notice that there are 2^{15} different "scenario types" for this problem since we consider three farm groups and five time periods.

Ethanol Demand & Candidate Biorefinery Locations

Blending facilities are petroleum refineries where the blending of biofuel and gasoline takes place. Thus, to calculate the biofuel demand, we first identify the existing petroleum refineries in Texas. As of 2012, there are 26 operational oil refineries in Texas and only 18 of them have a positive gasoline production capacity as presented in Section 2.6.3. Therefore, we use these 18 oil refineries as our demand locations. These locations are shown in Figure 4.2a. As this figure shows most of the blending facilities are located at the Gulf area.

The location and the gasoline production capacities of these refineries are obtained from the Annual Refinery Report of EIA [63]. We assume 50% of the total gasoline production capacity in these refineries utilized to satisfy demand in Texas with E-10 (10% Ethanol) gasoline-ethanol production. We obtain the ethanol upper bound demand in each blending facilities using these input values. Moreover, the same demand upper bound values are used for each time period. Our calculations show that the annual total ethanol upper bound demand is around 710 million

gallons.



(a) Blending facility locations (b) Candidate biorefinery locations

Figure 4.2: Demand and candidate biorefinery locations

To determine the candidate biorefinery locations, we use the data presented by EPA [67]. We use the same locations presented in Section 2.6.4, which are suitable for biorefinery construction with an "excellent" or "outstanding" potential rating and are at least 60 miles apart from each other. Figure 4.2b shows all the selected candidate biorefinery locations.

In our analysis, we take the holding and transportation costs for biofuel as 0.05 \$/gallon/year and 0.001 \$/gallon/mile, respectively, as suggested by Eksioglu et al. [18]. Moreover, we estimate the biofuel price as \$2.80/gallon for each time period in our base setting calculations. However, in Section 4.5.2, we alter the unit biofuel price and observe the changes in the supply chain system.

4.5.2 Analysis on Biofuel Prices

We solve the problem for different biofuel prices changing between \$2.60/gallon and \$3.00/gallon, and observe for which biofuel prices investing in biofuel production is worthwhile. We assume biofuel price is the same for all time periods. In total, we solve 21 problems with different biofuel prices, each problem having 15 instances for both SAAup and SAAlw.

Our results show that when unit biofuel price is lower than \$2.62/gallon, the expected total system profit is negative. In other words, the investment is not economically justified for the biofuel producer. Thus, based on our calculations, we can claim that the unit biofuel price needs to be at least \$2.62/gallon for the biofuel producer to invest in biofuel production. In this case, the biofuel producer makes an expected profit of \$7.74 million in 5 year planning horizon.

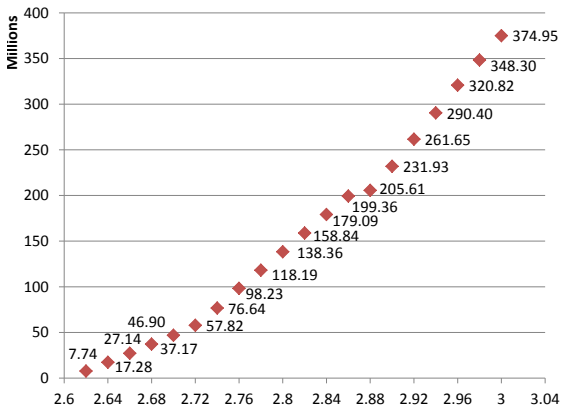


Figure 4.3: Biofuel price (\$/gallon) vs. Expected total system profit (\$)

As the unit biofuel price increases, the expected total system profit increases. For example, when the biofuel price is \$2.70/gallon, the expected system profit is around \$47 million. For higher biofuel prices, the expected system profit reaches

higher values. When biofuel price is $2.90/gallon for instance, the expected system profit is over $200 million. One interesting observation we make is that the expected system profit function resembles a piecewise linear function, which can be seen in Figure 4.3. We see three linear-like segments in Figure 4.3. The first linear segment is for biofuel prices between $2.64/gallon-$2.70/gallon. We observe that for biofuel prices between these values, the expected system profit increases around $10 million when the biofuel price increases by $0.02/gallon. For biofuel prices between $2.70/gallon and $2.86/gallon however, an increase of $0.02/gallon in biofuel price corresponds to an increase around $20 million in expected system profit. Lastly, for biofuel prices between $2.88/gallon and $3.00/gallon, we observe an increase around $30 million in the expected system profit, for increments of $0.02/gallon in biofuel price.

The unit biofuel price has a significant effect on the switchgrass wholesale price. Our results show that as biofuel price increases, the wholesale price offered to farmers also increases. For example, when biofuel price is $2.70/gallon, wholesale price is $38.54/dt. However, when biofuel price is $2.90/gallon, the optimal wholesale price is determined as $47.74/dt. This is because as the biofuel price increases, it becomes more profitable to process more switchgrass and produce more biofuel. Therefore, higher wholesale prices are offered to farmer to stimulate them to plant more switchgrass.

We see four major wholesale price options selected for different biofuel prices as it is seen in Figure 4.4. For biofuel prices between $2.62/gallon and $2.72/gallon, the optimal switchgrass wholesale price is $38.54/dt. When biofuel price is in between $2.74/gallon and $2.82/gallon, the optimal wholesale price offered to the farmers is $45.45/dt. Similarly, for biofuel price values in $2.82/gallon-$2.92/gallon, the wholesale price is $47.74/dt. Lastly, for biofuel prices larger than $2.92/gallon, the optimal

switchgrass wholesale price offered to farmers is determined as \$49.43/dt. Notice that when biofuel price is \$2.72/gallon, wholesale price is determined as \$42.97/dt, which is different than the four major biomass prices obtained.

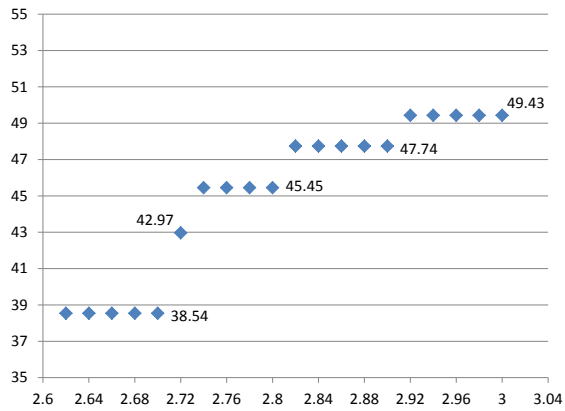


Figure 4.4: Biofuel price (\$/gallon) vs. Biomass price (\$/dt)

The network structure also changes as the biofuel price alters. For all biofuel prices, we observe that all the opened biorefineries are in large sizes. Moreover, we observe that the number of opened biorefineries increases as the unit biofuel price increases. This is because for higher biofuel prices more biofuel is produced to obtain more profit. We notice that for each of the four major wholesale prices shown in Figure 4.4, there is a corresponding unique network structure. Our results indicate that for biofuel prices between \$2.62/gallon and \$2.74/gallon, corresponding to wholesale prices \$38.54/dt and \$42.97/dt, only one biorefinery is opened at the location shown in Figure 4.5a. On the other hand, for biofuel prices between \$2.74/gallon and \$2.92/gallon, two biorefineries are opened as in Figure 4.5b and Figure 4.5c. Recall that for these biofuel prices, the optimal wholesale price is determined as \$45.45/dt and \$47.74/dt. Note that the biorefinery locations are different when biofuel price

is in \$2.74/gallon-\$2.82/gallon, and in \$2.82/gallon-\$2.92/gallon corresponding to optimal wholesale prices of \$45.45/dt and \$47.74/dt, respectively. Lastly, when the biofuel price is greater than \$2.92/gallon, we observe that three biorefineries are opened in locations shown in Figure 4.5d.

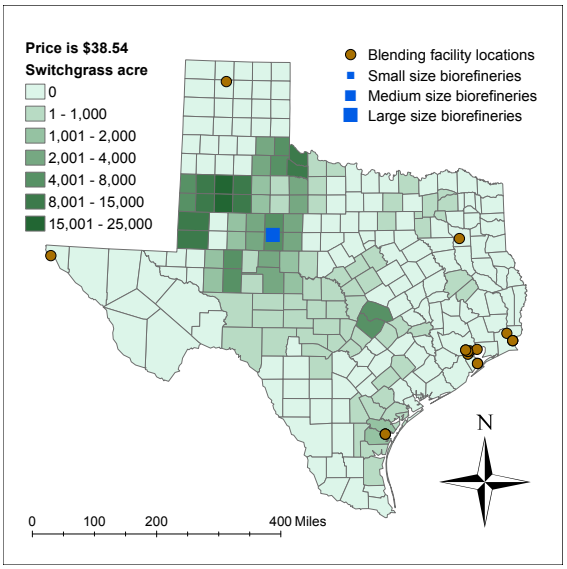
As mentioned, we observe a correlation between the wholesale price offered to farmers and the opened biorefinery locations. When there is only one biorefinery in the system, the optimal switchgrass wholesale price is determined as \$38.54/dt. Only when the biofuel price is \$2.72/gallon, there is one biorefinery opened and the wholesale price is \$42.97/dt. When there are two biorefineries in the supply chain network, the optimal wholesale price is either \$45.45/dt or \$47.74/dt. Lastly, when there are three biorefineries opened in the system, the optimal wholesale price offered to farmers is obtained as \$49.93/dt.

In our results, we see that the biofuel production investment is economically justifiable, and the biofuel producer makes profit in the 5 year planning horizon. It is interesting to see the pattern of switchgrass price and biorefinery locations changing according to biofuel price. This shows us the relationship and the connection between the supply chain network structure and biomass-biofuel prices.

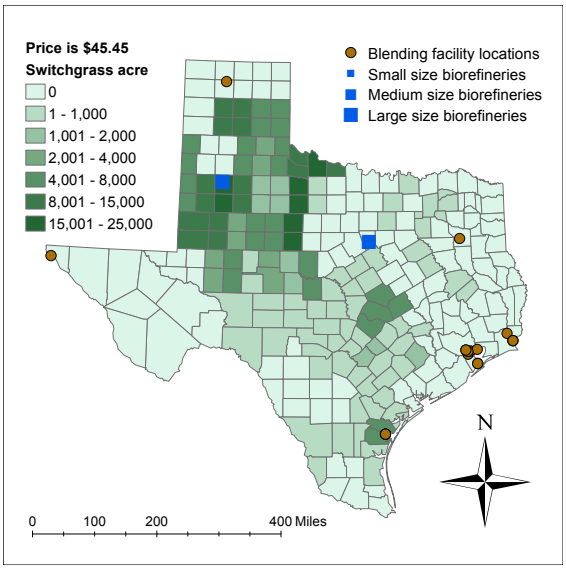
4.6 Concluding Remarks

In this study, we present a two-stage stochastic program to model a multi-period bio-energy supply chain system. Our model incorporates a wholesale price agreement between the biofuel producer and the farmers. Moreover, it takes the effect of switchgrass yield uncertainty and switchgrass price into account when calculating the switchgrass supply.

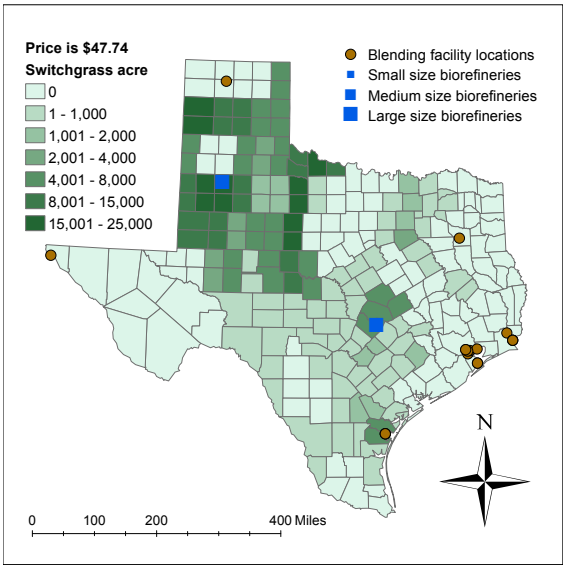
We implement an L-shaped method along with an SAA approach to solve this problem in a reasonable time and obtain statistically good solutions. Due to its



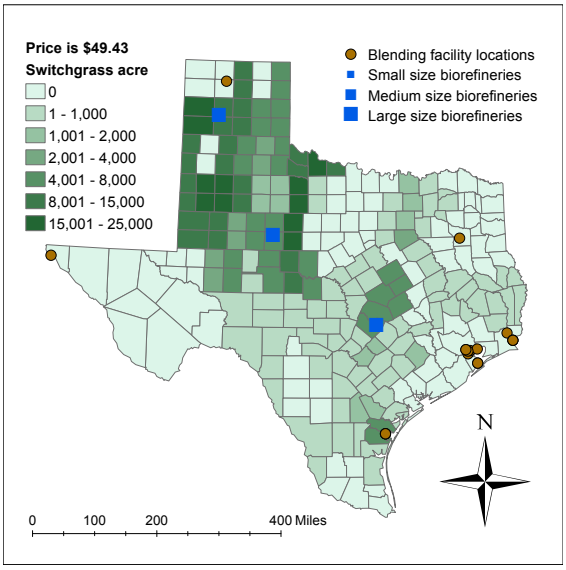
(a) Biofuel prices between \$2.62/gallon-\$2.72/gallon



(b) Biofuel prices between \$2.74/gallon-\$2.82/gallon



(c) Biofuel prices between \$2.82/gallon-\$2.92/gallon



(d) Biofuel prices between \$2.92/gallon-\$3.00/gallon

Figure 4.5: Supply chain network structure for different biofuel prices

multi-period nature, the problem shows significant uncertainty over the planning horizon and thus, we need to employ a large set of scenarios, i.e., 500 and 10,000 scenarios for the upper and lower bound SAA problems, while implementing the SAA approach to obtain statistically justifiable solutions.

In addition, we conduct a case study in Texas using realistic data utilized with GIS. Our results indicate that opening a biorefinery and producing biofuel is justifiable only when the biofuel price is higher than \$2.62/gallon. Making this investment for lower biofuel prices is economically not justified according to our results. Moreover, we investigate the relationship between biofuel price and switchgrass price. We observe that as the biofuel price increases, switchgrass price offered to the farmers also increases to cultivate more land and increase the expected switchgrass supply. Furthermore, the relationship between the supply chain network structure and the biofuel price is also explored. Our results indicate that more biorefineries are opened to produce and sell more biofuel as the biofuel price increases.

The tool developed in this study is helpful for biofuel producers who are interested in making investment in bio-energy. This tool shows whether the investment in biofuel production is profitable under a wholesale price commitment between a biofuel producer and a group of farmers. This tool assists the decision maker, i.e., the biofuel producer, to implement the right policies associated with its supply chain system. These policies include setting the correct switchgrass wholesale price and selecting the correct locations to open biorefineries for biofuel production. After implementing these optimal policies, the biofuel producer's profit is maximized over the planning horizon.

5. CONCLUSION

5.1 Summary

In this dissertation, we present three different modeling frameworks that design biomass-biofuel supply chains. Below, we summarize these studies and our results.

In Section 2, we present a deterministic multi-period and multi-biomass network design model. Our model determines both strategic and tactical level decisions associated with the bio-energy supply chain by integrating upstream and downstream echelons. We present a Benders Decomposition based solution algorithm that can efficiently solve large-scale problem instances, which we showed in our extensive computational study. Moreover, we conduct a case study in Texas utilizing GIS and using realistic data from the literature to demonstrate the capabilities of our model. We examine the changes in the supply chain network structure with respect to changes in some of the model input parameters such as conversion rates, transportation discount factors and supply and demand variability. Our results indicate that while the collection facility locations are quite sensitive to changes in input parameters, the biorefinery locations are relatively robust. We also observe that an integrated system approach in terms of biomass types and geographical area yields better solutions (8% cost reduction) than a segregated approach.

In our second study presented in Section 3, we propose a two-stage stochastic program to model farm-to-biorefinery biomass supply chain. We introduce a policy based on a wholesale price agreement between a group of farmers and a biofuel producer, to stimulate energy crop production. This policy considers farmer's decision making process related with biomass energy crop price and acreage. Our modeling framework is to simultaneously determine the policy parameter (biomass wholesale

price) and the biomass supply chain network structure. We implement a solution methodology that includes an algorithm based on L-shaped method to decompose the problem into smaller subproblems and an SAA approach to statistically justify our solutions. Our model and solution procedure are tested on a case study conducted in Texas utilizing GIS and realistic data from the literature. For different model input parameters, we observe the changes in the resulting policy parameter, expected total system cost and supply chain network structure. Our results indicate that changes in unit penalty and unit transportation costs considerably affect these results. Moreover, we monitor how the network structure and the expected total system cost change when we alter farmers' decision making parameters. We observe that the expected total system cost decreases when farmers are more optimistic, i.e., risk-taker, about energy crop yields.

Our last study presented in Section 4 extends the previous study and includes biofuel logistics in addition to biomass logistics. Similar to the previous model presented in Section 3, this model also utilizes a similar wholesale price agreement policy between farmers and biofuel producer. However, unlike the previous model, this model considers biofuel inventory decisions in a multi-period problem setting. An L-shaped method along with an SAA approach are implemented to solve this problem in a reasonable time and obtain statistically justifiable solutions. In addition, we conduct a case study in Texas using realistic data utilized in GIS. Our results show that opening a biorefinery and producing biofuel, i.e., investing in biofuel production, are economically justifiable only when biofuel price is higher than \$2.62/gallon. Moreover, as biofuel price increases, the policy parameter (switchgrass wholesale price) also increases to cultivate more land to energy crops and increase the expected switchgrass supply. Furthermore, we investigate the relationship between supply chain network structure and biofuel price. Our results indicate that there is a

correlation between switchgrass wholesale price and supply chain network structure.

5.2 Expected Significance of the Research

This research provides tools to effectively design biomass-biofuel supply chain network by considering its important and realistic characteristics. These tools presented in Sections 2, 3 and 4 provide assistance to biofuel industry in order to implement correct policies to reduce its supply chain costs and to increase its profitability and market share. The expected significance of this research can be grouped into four categories: (i) environmental, (ii) social and economic, (iii) theoretical and methodological, and (iv) application and case study.

On the environmental side, this research supports the good cause of using bio-energy and biofuels, which are environmental-friendly. Addressing and solving bio-energy supply chain issues help reducing bio-energy price and increase its accessibility and wide-range use. This makes societies less dependent on fossil fuels and results in less consumption of carbon based energies, which pollute our environment. In fact, using ethanol in place of gasoline helps to reduce carbon dioxide (CO2) emissions by an average of 34%. Moreover, ethanol also reduces tailpipe carbon monoxide emissions by as much as 30%, toxics content by 13% (mass) and 21% (potency), and tailpipe fine particulate matter (PM) emissions by 50% [43].

On the social and economic side, bio-energy and biofuel industry provide job creation along their supply chain. This boosts the economy as well as social life and creates new opportunities for societies. In fact, in 2013, the U.S. ethanol industry added $44 billion to the U.S. Gross Domestic Product (GDP) and paid $8.3 billion in taxes. Moreover, in the same year, ethanol industry supported 86,503 direct jobs in renewable fuel production and agriculture sectors in the U.S., as well as 300,279 indirect and induced jobs in the overall economy [69]. The contributions of biofuel

industry to the economy will continue with an increasing pace in the future. The U.S. Department of Energy predicts that 10,000 to 20,000 jobs will be added to the economy for every one billion gallons of ethanol produced [43].

On the theoretical and methodological side, this research provides (i) mathematical models to design biomass-biofuel supply chain networks to optimally determine the strategic and/or tactical decisions considering important biomass-biofuel supply chain features, (ii) policies between biomass-biofuel supply chain actors to increase biomass supply, especially for second-generation (advanced) biofuel production, (iii) insights on the relationship between biomass price and biomass supply chain network structure as well as their effects on each other, (iv) solution frameworks including algorithmic approaches based on decomposition methods that are capable of handling large-scale problems.

Lastly, this research includes extensive case studies conducted in Texas that has very rich natural and economic resources. Real data related to biomass-biofuel supply chain, is gathered from the literature utilizing GIS and employed to obtain realistic results. Moreover, sensitivity analysis on many model input parameters is performed to observe system dynamics. Our case study results provide insights on the bioenergy and biofuel potential in Texas and might be beneficial for parties that plan to invest in biofuel industry. It is our belief that more biofuel investment in Texas will take place eventually since Texas is one of the largest ethanol-consuming states. In fact, in 2011, ethanol consumption in Texas constitutes 31% of the total ethanol consumption in the U.S. [61].

5.3 Future Research Directions

In a recent study, Sharma et al. [49] identify five future research directions for biomass-biofuel supply chain design. These directions are (i) incorporating uncertainty into model formulations, (ii) focusing on system wide (integrated) modeling and optimization, (iii) developing models that can be adapted by specific stakeholders such as farmers and biofuel producers, (iv) addressing computational complexities and developing large-scale case studies, (v) incorporating economic, social and environmental measures into model formulations. The scope of this dissertation covers the first four points that Sharma et al. [49] identify. However, the last point is not addressed in this research.

In their study, De Meyer et al. [14] state that the current literature on biomass-biofuel supply chain lacks of mathematical models that consider economical, environmental and social objectives in an integrated fashion. Shabani et al. [47] also agree to this statement and argue that the future research in bio-energy supply chain should include environmental and social impacts of bio-energy such as carbon emissions, job creation, governmental policies etc., in addition to the economic objectives.

Therefore, we believe that future research areas should incorporate the economic, social and environmental measures in an integrated fashion into biomass-biofuel supply chain models. This way the trade-offs involving not only the economic but social and environmental issues are addressed and a more comprehensive evaluation of bio-energy can be performed.

REFERENCES

[1] H. An, W. E. Wilhelm, and S. W. Searcy. A mathematical model to design a lignocellulosic biofuel supply chain system with a case study based on a region in Central Texas. *Bioresource Technology*, 102:7860–7870, 2011.

[2] H. An, W. E. Wilhelm, and S. W. Searcy. Biofuel and petroleum-based fuel supply chain research: A literature review. *Biomass & Bioenergy*, 35(9):3763–3774, 2011.

[3] R. C. Badger. Processing cost analysis for biomass feedstocks. Technical report, Oak Ridge National Laboratory, October 2002. ORNL/TM-2002/199.

[4] Y. Bai, Y. Ouyang, and J. Pang. Biofuel supply chain design under competitive agricultural land use and feedstock market equilibrium. *Energy Economics*, 34 (5):1623–1633, 2012.

[5] S. Banerjee, S. Mudliar, R. Sen, B. Giri, D. Satpute, T. Chakrabarti, and R. A. Pandey. Commercializing lignocellulosic bioethanol: Technology bottlenecks and possible remedies. *Biofuels Bioproducts & Biorefining-Biofpr*, 4(1):77–93, 2010.

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

[7] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer, New York, NY, 1997.

[8] G. Bocqueho and F. Jacquet. The adoption of switchgrass and miscanthus by farmers: Impact of liquidity constraints and risk preferences. *Energy Policy*, 38 (5):2598–2607, 2010.

[9] I. M. Bowling, J. M. Ponce-Ortega, and M. M. El-Halwagi. Facility location and

supply chain optimization for a biorefinery. *Industrial & Engineering Chemistry Research*, 50(10):6276–6286, 2011.

[10] C. Chen and Y. Fan. Bioethanol supply chain system planning under supply and demand uncertainties. *Transportation Research Part E - Logistics and Transportation Review*, 48(1, SI):150–164, 2012.

[11] X. Chen and H. Onal. An economic analysis of the future U.S. biofuel industry, facility location, and supply chain network. http://dx.doi.org/10.1287/trsc.2013.0488, 2013. Forthcoming in *Transportation Science*.

[12] M. Dal-Mas, S. Giarola, A. Zamboni, and F. Bezzo. Strategic design and investment capacity planning of the ethanol supply chain under price uncertainty. *Biomass & Bioenergy*, 35(5):2059–2071, 2011.

[13] M. Daroch, S. Geng, and Wang G. Recent advances in liquid biofuel production from algal feedstocks. *Applied Energy*, 102:1371–1381, 2013.

[14] A. De Meyer, D. Cattrysse, J. Rasinmaeki, and J. Van Orshoven. Methods to optimise the design and management of biomass-for-bioenergy supply chains: A review. *Renewable & Sustainable Energy Reviews*, 31:657–670, 2014.

[15] A. Demirbas. Biomass resource facilities and biomass conversion processing for fuels and chemicals. *Energy Conversion And Management*, 42(11):1357–1378, 2001.

[16] A. Demirbas. Political, economic and environmental impacts of biofuels: A review. *Applied Energy*, 86, Supplement 1:S108 – S117, 2009.

[17] M. Downing and R. L. Graham. The potential supply and cost of biomass from energy crops in the Tennessee valley authority region. *Biomass & Bioenergy*, 11 (4):283–303, 1996.

[18] S. D. Eksioglu, A. Acharya, L. E. Leightley, and S. Arora. Analyzing the design

and management of biomass-to-biorefinery supply chain. *Computers & Industrial Engineering*, 57:1342–1352, 2009.

[19] Ethanol Producer Magazine. USA plants. `http://www.ethanolproducer.com/plants/listplants/USA/`, 2012. Last viewed in 5/2012.

[20] O. E. Flippo and A. H. G. Rinnooy Kan. A note on Benders Decomposition in mixed integer quadratic-programming. *Operations Research Letters*, 9(2):81–83, 1990.

[21] Frontier Associates LLC. Texas renewable energy resource assessment 2008, December 2008.

[22] C. D. Garland. Growing and harvesting switchgrass for ethanol production in Tennessee. `https://utextension.tennessee.edu/publications/Documents/SP701-A.pdf`, 2010. Last viewed in 09/2014.

[23] A. M. Geoffrion and G. W. Graves. Multicommodity distribution system-design by Benders Decomposition. *Management Science Series A-Theory*, 20(5):822–844, 1974.

[24] J. Goldemberg. Ethanol for a sustainable energy future. *Science*, 315(5813): 808–810, 2007.

[25] C. N. Hamelinck, G. van Hooijdonk, and A. P. C. Faaij. Ethanol from lignocellulosic biomass: Techno-economic performance in short-, middle- and long-term. *Biomass & Bioenergy*, 28(4):384–410, 2005.

[26] Y. Huang, C. Chen, and Y. Fan. Multistage optimization of the supply chains of biofuels. *Transportation Research Part E - Logistics and Transportation Review*, 46(6):820–830, 2010.

[27] IBM. User manual for cplex v12.4. `http://pic.dhe.ibm.com/infocenter/cosinfoc/v12r4/index.jsp`, 2011. Last viewed in 10/2013.

[28] K. Jensen, C. D. Clark, P. Ellis, B. English, J. Menard, M. Walsh, and D. Ugarte.

Farmer willingness to grow switchgrass for energy production. *Biomass & Bioenergy*, 31(11-12):773–781, 2007.

[29] J. Kim, M. J. Realff, and J. H. Lee. Optimal design and global sensitivity analysis of biomass supply chain networks for biofuels under uncertainty. *Computers & Chemical Engineering*, 35(9, SI):1738–1751, 2011.

[30] A.J. Kleywegt, A. Shapiro, and T. Homem-De-Mello. The sample average approximation method for stochastic discrete optimization. *Siam Journal On Optimization*, 12(2):479–502, 2001.

[31] P. D. Kofman. Quality wood chip fuel. COFORD Connects Notes: Harvesting/Transportation No.6, 2006.

[32] J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142(1): 215–241, 2006.

[33] B. Lippke, M. E. Puettmann, L. Johnson, R. Gustafson, R. Venditti, P. Steele, J. F. Katers, A. Taylor, T. A. Volk, E. Oneil, K. Skog, E. Budsberg, J. Daystar, and J. Caputo. Carbon emission reduction impacts from alternative biofuels. *Forest Products Journal*, 62(4), 2012.

[34] T. L. Magnanti and R. T. Wong. Accelerating Benders Decomposition - Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3): 464–484, 1981.

[35] W. A. Marvin, L. D. Schmidt, and P. Daoutidis. Biorefinery location and technology selection through supply chain optimization. *Industrial & Engineering Chemistry Research*, 52(9):3192–3208, 2013.

[36] K. McNew and D. Griffith. Measuring the impact of ethanol plants on local grain prices. *Review Of Agricultural Economics*, 27(2):164–180, 2005.

[37] A. Mondala, K. Liang, H. Toghiani, R. Hernandez, and T. French. Biodiesel

production by in situ transesterification of municipal primary and secondary sludges. *Bioresource Technology*, 100(3):1203–1210, 2009.

[38] National Renewable Energy Laboratory. NREL 2008 biomass data. `http://www.nrel.gov/gis/biomass`, 2008. Last viewed in 5/2012.

[39] A. Okwo and V. M. Thomas. Biomass feedstock contracts: Role of land quality and yield variability in near term feasibility. *Energy Economics*, 42:67–80, 2014.

[40] A. Osmani and J. Zhang. Stochastic optimization of a multi-feedstock lignocellulosic-based bioethanol supply chain under multiple uncertainties. *Energy*, 59:157–172, 2013.

[41] A. Osmani and J. Zhang. Economic and environmental optimization of a large scale sustainable dual feedstock lignocellulosic-based bioethanol supply chain in a stochastic environment. *Applied Energy*, 114(SI):572–587, 2014.

[42] D. Rajagopal, S. Sexton, G. Hochman, D. Roland-Holst, and D. Zilberman. Model estimates food-versus-biofuel trade-off. *California Agriculture*, 63(4):199–201, 2009.

[43] Renewable Fuels Association. Ethanol facts: Environment. `http://www.ethanolrfa.org/pages/ethanol-facts-environment`, 2014. Last viewed in 09/2014.

[44] M. W. Rosegrant, S. Msangi, T. Sulser, and R. Valmonte-Santos. Biofuels and global food balance. *Bioenergy Andagriculture: Promises And Challenges*, 14 (3), 2006.

[45] P. Rubin. Benders Decomposition then and now. `http://orinanobworld.blogspot.com/2011/10/benders-decomposition-then-and-now.html`, 2011. Last viewed in 08/2013.

[46] R. Schnepf. Cellulosic ethanol: Feedstocks, conversion technologies, economics, and policy options. Technical report, Congressional Research Service, 2010.

R41460.

[47] N. Shabani, S. Akhtari, and T. Sowlati. Value chain optimization of forest biomass for bioenergy production: A review. *Renewable & Sustainable Energy Reviews*, 23:299–311, 2013.

[48] B. Sharma, R. G. Ingalls, C. L. Jones, R. L. Huhnke, and A. Khanchi. Scenario optimization modeling approach for design and management of biomass-to-biorefinery supply chain system. *Bioresource Technology*, 150:163–171, 2013.

[49] B. Sharma, R. G. Ingalls, C. L. Jones, and A. Khanchi. Biomass supply chain design and analysis: Basis, overview, modeling, challenges, and future. *Renewable & Sustainable Energy Reviews*, 24:608–627, 2013.

[50] R. E. H. Sims, W. Mabee, J. N. Saddler, and M. Taylor. An overview of second generation biofuel technologies. *Bioresource Technology*, 101(6):1570–1580, 2010.

[51] D. H. S. Tay, D. K. S. Ng, and R. R. Tan. Robust optimization approach for synthesis of integrated biorefineries with supply and demand uncertainties. *Environmental Progress & Sustainable Energy*, 32(2):384–389, 2013.

[52] Texas A&M Agrilife Extension Service. Extension agricultural economics. `http://agecoext.tamu.edu/resources/crop-livestock-budgets/budgets-by-commodity/`. Last viewed in 3/2014.

[53] Texas Department of Transportation. Regional service centers. `http://www.dot.state.tx.us/local_information/regions/default.htm`, 2012. Last viewed in 5/2012.

[54] United States Department of Agriculture. USDA national agricultural statistics service. `http://quickstats.nass.usda.gov/`, . Last viewed in 5/2013.

[55] United States Department of Agriculture. Texas agricultural statistical districts. `http://www.nass.usda.gov/Statistics_by_State/Texas/Charts_&_`

Maps/distmap2.htm, . Last viewed in 3/2014.

[56] United States Department of Agriculture. USDA 2007 census of agriculture. http://www.agcensus.usda.gov/Publications/2007/Full_Report/Volume_1,_Chapter_2_County_Level/Texas/, 2007.

[57] United States Department of Agriculture. USDA biofuels strategic production report. http://www.usda.gov/documents/USDA_Biofuels_Report_6232010.pdf, June 2010.

[58] United States Department of Agriculture. Crop values 2013 summary. http://usda01.library.cornell.edu/usda/current/CropValuSu/CropValuSu-02-14-2014.pdf, February 2014.

[59] United States Department of Energy. Theoretical ethanol yield calculator. http://www1.eere.energy.gov/biomass/ethanol_yield_calculator.html, . Last viewed in 5/2012.

[60] United States Department of Energy. Biomass feedstock composition and property database. http://www.afdc.energy.gov/biomass/progs/search1.cgi, . Last viewed in 5/2012.

[61] United States Department of Energy. Energy consumption by transportation fuel in texas. http://apps1.eere.energy.gov/states/transportation.cfm/state=TX, 2011.

[62] United States Energy Information Administration. EIA projects world energy use to increase 53 percent by 2035; china and india account for half of the total growth. http://www.eia.gov/pressroom/releases/press368.cfm, 2011. Last viewed in 09/2014.

[63] United States Energy Information Administration. Refinery capacity report. http://www.eia.gov/petroleum/refinerycapacity/, 2012. Last viewed in 5/2012.

[64] United States Energy Information Administration. Annual energy review. http://www.eia.gov/totalenergy/data/annual/pecss_diagram.cfm, September 2012. Last viewed on 09/2014.

[65] United States Environmental Protection Agency. Renewable fuel standard (RFS). http://www.epa.gov/OTAQ/fuels/renewablefuels/. Last viewed on 09/2014.

[66] United States Environmental Protection Agency. Biomass CHP catalog. http://www.epa.gov/chp/documents/biomass_chp_catalog_part3.pdf, 2007.

[67] United States Environmental Protection Agency. EPA tracked sites with clean and renewable energy generation potential. http://www.epa.gov/renewableenergyland/data.htm, 2012. Last viewed in 5/2012.

[68] United States Internal Revenue Services. Publication 946, how to depreciate propoerty. http://www.irs.gov/pub/irs-pdf/p946.pdf, January 2014.

[69] J. M. Urbanchuk. Contribution of the ethanol industry to the economy of the United States. Technical report, Renewable Fuels Association, 2014.

[70] H. Üster and P. Kewcharoenwong. Strategic design and analysis of a relay network in truckload transportation. *Transportation Science*, 45(4):505–523, 2011.

[71] M. B. Villamil, M. Alexander, A. H. Silvis, and M. E. Gray. Producer perceptions and information needs regarding their adoption of bioenergy crops. *Renewable & Sustainable Energy Reviews*, 16(6):3604–3612, 2012.

[72] R. Wallace, K. Ibsen, A. McAloon, and W. Yee. Feasibility study for co-locating and integrating ethanol production plants from corn starch and lignocellulosic feedstocks. Technical report, National Renewable Energy Laboratory, 2005. http://www1.eere.energy.gov/biomass/pdfs/37092.pdf.

[73] D. Yue, F. You, and S. W. Snyder. Biomass-to-bioenergy and biofuel supply

chain optimization: Overview, key issues and challenges. *Computers & Chemical Engineering*, 66:36–56, 2014.

[74] J. Zhang, A. Osmani, I. Awudu, and V. Gonela. An integrated optimization model for switchgrass-based bioethanol supply chain. *Applied Energy*, 102:1205–1217, 2013.

[75] X. Zhu and Q. Yao. Logistics system design for biomass-to-bioenergy industry with multiple types of feedstocks. *Bioresource Technology*, 102(23):10936–10945, 2011.

[76] X. Zhu, X. Li, Q. Yao, and Y. Chen. Challenges and models in supporting logistics system design for dedicated-biomass-based bioenergy industry. *Bioresource Technology*, 102(2):1344–1351, 2011.