Models and algorithms to solve a reliable and congested biomass supply chain network designing problem under uncertainty.

By

Sushil Raj Poudel

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Industrial and Systems Engineering
in the Department of Industrial and Systems Engineering

Mississippi State, Mississippi

May 2017

Models and algorithms to solve a reliable and congested biomass supply chain network designing problem under uncertainty.

By

Sushil Raj Poudel

Approved:

Mohammad Marufuzzaman
(Major Professor)

Linkan Bian
(Co-Major Professor)

Hugh R. Medal
(Committee Member)

Li Zhang
(Committee Member)

Reuben F. Burch V
(Committee Member)

Stanley F. Bullington
(Graduate Coordinator)

Jason M. Keith
Dean
Bagley College of Engineering

Name: Sushil Raj Poudel

Date of Degree: May 5, 2017

Institution: Mississippi State University

Major Field: Industrial and Systems Engineering

Major Professor: Dr. Mohammad Marufuzzaman

Title of Study: Models and algorithms to solve a reliable and congested biomass supply chain network designing problem under uncertainty.

Pages of Study: 175

Candidate for Degree of Doctor of Philosophy

This dissertation studies two important problems in the field of biomass supply chain network. In the first part of the dissertation, we study the pre-disaster planning problem that seeks to strengthen the links between the multi-modal facilities of a biomass supply chain network. A mixed-integer nonlinear programming model is developed to determine the optimal locations for multi-modal facilities and bio-refineries, offer suggestions on reliability improvement at vulnerable links, production at bio-refineries, and make transportation decision under both normal and disrupted scenarios. The aim is to assist investors in determining which links' reliability can be improved under specific budget limitations so that the bio-fuel supply chain network can prevent possible losses when transportation links are disrupted because of natural disasters. We used states Mississippi and Alabama as a testing ground for our model. As part of numerical experimentation, some realistic hurricane scenarios are presented to determine the potential impact that pre-investing may have

on improving the bio-mass supply chain network's reliability on vulnerable transportation links considering limited budget availability.

In the second part of the dissertation, we study the impact of feedstock supply uncertainty on the design and management of an inbound biomass co-firing supply chain network. A two-stage stochastic mixed integer linear programming model is developed to determine the optimal use of multi-modal facilities, biomass storage and processing plants, and shipment routes for delivering biomass to coal plants under feedstock supply uncertainty while considering congestion into account. To represent a more realistic case, we generated a scenario tree based on the prediction errors obtained from historical and forecasted feedstock supply availability. We linearized the nonlinear problem and solved with high quality and in a time efficient manner by using a hybrid decomposition algorithm that connects a Constraint generation algorithm with Sample average approximation algorithm and enhanced Progressive hedging algorithm. We used states Mississippi and Alabama as a testing ground for our study and conducted thorough computational experiments to test our model and to draw managerial insights.

Key words: Biomass supply chain network, disaster management, network reliability, congestion prevention, sample average approximation, progressive hedging algorithm, rolling horizon heuristics

DEDICATION

To my parents and grandparents.

ACKNOWLEDGEMENTS

I am truly indebted and thankful to my advisor, Dr. Mohammad Marufuzzaman, for introducing me to the scientific research. He always provided me the freedom to explore and proper guidelines. Moreover, his patience and trust towards me have been a great motivation throughout these four years. I consider it an honor to work with my co-major advisor Dr. Linkan Bian and my committee members Dr.Hugh Medal, Dr.Reuben F. Burch V,and Dr. Li Zhang.

I would especially like to thank my lab-mates Sudipta, Rasel, Amir, MK and Ravi for all those fruitful discussions on academic and casual matters. I also thank all other friends and colleagues at McCain building and Starkville for not keeping me isolated and lonely. I also want to thank Holmes Cultural Diversity Center(HCDC), FedEx and Department of industrial and systems engineering for their financial support. At the end, I would like to express my appreciation to my sister Shrijana, brother-in-law Suresh, brother Sanjaya, dad Dhruba, mum Shanti and girl friend Dipa, whose support, encouragement, and unwavering love on all high and lows gave the the power to proceed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Energy products and transportation fuels produced from the use of biomass are proposed as the solution of our dependence on fossil fuels and increasing pollution of environment because of green house gases (GHG) emission. Biomass is renewable source of energy and the process of production of bioenergy/biofuel from biomass along with the emission during the use of biofuel have lower impact on environment in comparison to the production and use of fossil fuels. The Energy Policy Act of 2005 and the Energy Independence and Security Act of 2007 [114] mandate the nation's annual bio-ethanol production should grow from 10.01 billion gallons in 2011 to over 36 billion gallons by 2022 [115]. Similarly, many countries throughout the world have set their targets on biofuel uses and provide incentive and benefits to industries for promoting the growth of biomass uses for energy production. Bioenergy industry will be undergoing a rapid expansion on the coming decade and such an increase in bio-energy production highlights the need for developing a reliable and efficient supply chain system that not only performs well under normal conditions but also hedges against risk under various unexpected disruption scenarios, uncertainties and possible congestion on network and facilities because of overflow of the commodities.

After a series of devastating disasters in recent years, including the U.S. Northeast blackout in 2003, Hurricane Katrina in 2005, earthquakes in China in 2008, Haiti in 2009, and Nepal in 2015 and the Gulf of Mexico oil spill in 2010, it is evident that the transportation infrastructure, particularly those bearing multi-modal traffic, is vulnerable to various disruption risks. If the transportation link fails, several weeks, months, or years may be required to restore it back to the normal condition. Therefore, strengthening the weakest components of the transportation links is necessary to enhance their survivability even after hurricanes strike. In Chapter 1 we provide the optimization model with the objective of maximizing the post-disaster connectivity of multi-modal facility links while minimizing the investment and transportation costs among multiple sources to their respective destination nodes in a bio-fuel supply chain network.

Co-firing is a low-cost option to convert biomass to electricity in efficient manner and in very clean way by adding biomass as a substitute feeder on a coal plant boilers. There is no loss on the efficiency of boilers by co-firing, biomass combustion efficiency to electricity will be close to 33%-37% while co-firing [85]. The use of biomass for electricity production is expected to increase. It has great opportunity for co-firing since the coal powered boilers where co-firing is possible has capacity of generating 310 gigawatts of electricity.

However, the uncertainty in the availability of biomass is major challenge in long-term strategic planning of biomass supply system. Biomass supply face more uncertainties compared with the supply of coal due to unpredictable weather condition, and natural disaster like hurricane and tornadoes. The effects of these uncertainities affect's the efficiency of the

system. It will lead either to an in-feasible supply chain network designs or to sub-optimal performance. Uncertainty and seasonality of the biomass availability can also lead to congestion on the multi-modal facilities and links. Congestion causes delays and it will affect the reliability of the business, its schedule, production, demand fulfillment and many more. It not only affect the congested zone, it will have a chain effect throughout the enterprise. Designing a network considering the congestion and its possible impacts and assigning the flows in an optimal way while preventing overuse of a hub to cause congestion is required. On the ongoing work, we had developed a comprehensive stochastic program model for the optimal design and planning of biomass co-firing supply chain that is robust from the economic performance perspective under the presence of supply uncertainties and prevent congestion from occurring or minimize the effect of congestion.

In Chapter 2, we present a pre-disaster planning model that seeks to strengthen a bio-fuel supply chain system's multi-modal facility links while accounting for limited budget availability. The proposed model determines which set of facilities and links to select that will maximize post-disaster connectivity and minimize bio-fuel supply chain related costs. The failure probability of the links between the multi-modal facilities is estimated using a spatial statistic model, which is developed from real world data. We solve this challenging $\mathcal{NP}$-hard problem developing a generalized Benders decomposition algorithm. The proposed algorithm is validated via a real-world case study with data from Mississippi and Alabama. Computational results show that the proposed solution approach is capable of solving the problem efficiently. Several experiments are conducted to demonstrate the applicability of this model by testing various model parameters on bio-fuel supply

chain network performance, including reliability improvement cost, availability of budget, biomass supply changes, and the risk averseness degree for decision makers. Numerical analysis indicates that, under normal conditions, the minimum cost model determines a unit bio-fuel delivery cost of $3.56/gallon. However, in case of a disaster, the unit bio-fuel delivery cost provided by the minimum cost model increases to $3.96/gallon, compared to $3.69/gallon provided by the reliable model solution.

In Chapter 3, we present a two-stage stochastic programming model for the design and management of a biomass co-firing supply chain network under feedstock supply uncertainty. To represent a more realistic case, we generate scenarios from prediction errors of the historical and forecasted biomass supply availabilities. We solve the model using a hybrid decomposition algorithm that combines Sample average approximation with an enhanced Progressive hedging algorithm. The proposed algorithm is validated via a real-world case study using data from Mississippi and Alabama. Computational results indicate that the proposed algorithm is capable of producing high quality solutions in a reasonable amount of time.

In Chapter 4, we further extend the model presented in Chapter 3 by taking congestion effects into account. A non-linear cost term is added in the objective function representing the congestion factor which increases exponentially as flow of biomass approaches the capacity of multi-modal facility. We first linearize the model and then use a nested decomposition algorithm to obtain a feasible solution in a reasonable amount of time. The nested decomposition algorithm that we propose combine Constraint Generation (CG) algorithm with a Sample Average Approximation (SAA) and Progressive Hedging (PH) algorithm.

We apply some heuristics such as Rolling Horizon (RH) algorithm and variable fixing technique to enhance the performance of the PH algorithm. We develop a case study using data from the states of Mississippi and Alabama and use those regions to test and validate the performance of the proposed algorithm. Numerical experiments show that the proposed algorithm can solve large-scale problems with a larger number of scenarios and time periods to a near optimal solution in a reasonable amount of time. Results obtained from the experiments reveal that the delivery cost increases and less hubs with higher capacity are selected if we take congestion cost into account.

In summary, our study provides models and algorithms to solve a reliable and congested biomass supply chain network. We have conducted a number of real life experiments that will help decision makers to design a reliable inbound logistics network for biomass supply chain.

CHAPTER 2

DESIGNING A RELIABLE BIO-FUEL SUPPLY CHAIN NETWORK CONSIDERING LINK FAILURE PROBABILITIES

2.1 Introduction

The U.S. bio-fuel industry is expanding at a phenomenal rate because of the implementation of the Energy Policy Act of 2005 and the Energy Independence and Security Act of 2007 [114]. Both mandate the nation's annual bio-ethanol production should grow from 10.01 billion gallons in 2011 to over 36 billion gallons by 2022 [115]. Such an increase in bio-ethanol production highlights the need for developing a reliable and efficient supply chain system that not only performs well under normal conditions but also hedges against risk under various unexpected disruption scenarios.

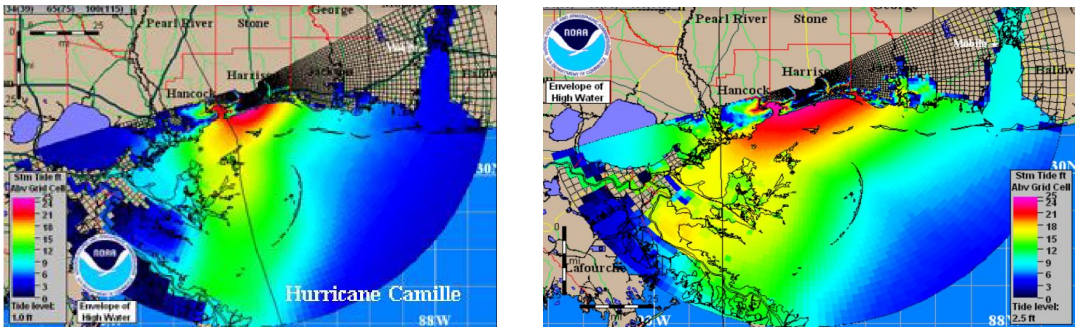
After a series of devastating disasters in recent years, including the U.S. Northeast blackout in 2003, Hurricane Katrina in 2005, earthquakes in China in 2008 and Haiti in 2009, and the Gulf of Mexico oil spill in 2010, it is evident that the transportation infrastructure, particularly those bearing intermodal traffic, is vulnerable to various disruption risks. Figure 2.1 shows the impacts of hurricanes Katrina and Camille on the coasts of Mississippi and Alabama; each color represents the intensity of the hurricanes' devastating impact on those regions. The figure suggests that the failure probability of transportation infrastructure decreases for routes located farther away from the hurricane center. In real

life cases, the multi-modal facilities located farther away from the hurricane center usually survive; however, the links that connect the multi-modal facilities may be easily disrupted due to several reasons such as bridge failure, flash flooding[1], landslides[2], or other unexpected events [89, 52]. If the transportation link fails, several weeks, months, or years may be required to restore it back to the normal condition. Therefore, strengthening the weakest components of the transportation links is necessary to enhance their survivability even after hurricanes strike. In an ideal case, the decision maker strengthen all the links to a targeted safety level, but because of the potentially high expenditure, a subset of the links should be selected to meet specified budget limitations. Thus, the optimization model presented here has the objective of maximizing the post-disaster connectivity of multi-modal facility links while minimizing the investment and transportation costs among multiple sources to their respective destination nodes in a bio-fuel supply chain network.

One major stream of research in bio-fuel supply chain literature identifies the optimal locations of bio-refineries to minimize the cost of the overall supply chain network. Such problems are generally considered as extensions of classical facility location design problem. The literature on the topic of facility location and supply chain design is very rich. The first work on facility location problem dates back to the Weber problem [125] and since then extensive research have been conducted on different variants of the classical facility location problem, such as Canel et al., [13], Shiode and Drezner [105], Carrizosa and

[1] A rapid rise in water levels that can occur quickly due to intense rainfall over a relatively short period of time

[2] A landslide is a geological phenomenon in which ground movements, such as rock falls, deep slope failures, and shallow debris flows, occur and are driven downward by gravity. Landslides are caused when the stability of a slope changes from stable to unstable condition. [52]



(a) Hurricane Camille (b) Hurricane Katrina

Figure 2.1

Impacts of hurricanes Katrina and Camille on the coasts of Mississippi and Alabama [124]

Nickel [15], Melo et al., [78], Farahani et al., [34], Lee and Dong [63], Torres-Soto and Uster [112], Jena et al., [55], and others. Interested readers are requested to check some most recent development of facility location problem from the review papers of Melkote and Daskin [77], Klose and Drexl [61], Snyder [106], and Melo et al. [79]. Studies conducted by Zamboni et al. [132], Eksioglu et al. [29], Huang et al. [49], Xie and Ouyang [129], Xie et al. [128], Marufuzzaman et al. [74], Meyer et al. [81], and Walther et al. [121] develop deterministic models optimizing both plant locations and transportation costs in bio-fuel supply chain networks. Chen and Fan [17], Kim et al. [59], and Marufuzzaman et al. [75] extend those formulations by incorporating stochastic variants in the optimization model. The aim is to generate reliable solutions for the design and management of a bio-fuel supply chain network. However, one of the major drawbacks of these studies is that the authors assume the supply chain network will always function perfectly, and it will never fail. In reality, supply chain networks may be potentially impacted by the unexpected failure of transportation links [20, 91].

Supply chain reliability and resilience against natural disasters have gained increasing attention. Studies conducted by Daskin [24, 25] first considers facility unavailability in a maximal covering location problem. Drezner [28] extend those works by developing a reliable p-median location problem. Snyder and Daskin [107] further extend those models to consider a reliable uncapacitated fixed-charge location problem (UFLP) and the p-median problem where the authors assumed that the facilities may fail randomly with identical probabilities. Over time Cui et al. [22], Li and Ouyang [65], Shen et. al [104], and Li et al. [64] extend the existing models by relaxing the uniform disruption probability assumption introduced by [107]. Studied conducted Galindo and Batta [36], Rawls and Turnquist [96, 98, 97] are more relevant to this study where the authors attempted to develop models for pre-positioning of emergency supplies to hedge against potential natural disasters. Most recently, some studies attempt to address the impact of bio-refinery disruption in a bio-fuel supply chain network. Li et al. [66] develop a discrete and a continuous model to design a reliable bio-fuel supply chain network. The authors consider both site-independent and dependent disruptions, and they analyze the impact of those disruption probabilities on optimal refinery deployment decisions. Wang and Ouyang [122] propose a game-theoretical based continuous approximation model to locate bio-refineries under spatial competition and facility disruption risks. A very recent study by Marufuzzaman et al. [76] considers the impact of intermodal hub disruption on bio-fuel supply chain network design. However, none of these previous studies account for the probability of failure along the multi-modal facility-to-facility links because of the challenges associated with collecting data about link failures. This paper considers a pre-disaster planning problem

that seeks to reduce or eliminate risks between the multi-modal facility-to-facility links subject to budget constraints. A spatial statistics model is utilized to estimate the probability of link failure based on disaster disruption data that is already available at finite locations, e.g., multi-modal facilities.

Another stream of research has been conducted to develop link improvement plans for better disaster response. These studies primarily focus on factors specific to the physical characteristics of the links and the cost to upgrade them to withstand disaster under a specific severity level [8, 108]. Moghtaderi-Zadeh and Der Kiureghian [83] propose an efficient procedure for upgrading the reliability of existing lifeline networks to maximize post-earthquake serviceability. The authors then use a heuristic approach to invest incrementally in critical components that increase the system reliability beyond a given target value. Outside the disaster context, several studies address the stochastic degradation of link capacity occurring because of day-to-day traffic incidents [68, 67]. Clark and Watling [18] consider the effect of stochastic demand on network reliability, whereas Lo et al. [67] investigate the scenario related to simultaneous consideration of stochastic demand and link capacity degradation. However, previous studies do not show how to make link investment decisions to increase reliability. To the best of our knowledge, the only relevant study is the paper by Peeta et al. [89] which considers a pre-disaster planning problem in disaster management that facilitate highway network investment decisions to enhance disaster response. However, the major difference between their work and this paper is that this work considers multi-modal facility-link improvement investment decisions in the context of designing a supply chain network.

To the best of the authors' knowledge, the impact of multi-modal facility link failure and its impact on the bio-fuel supply chain network has not been thoroughly studied. Specifically, the need for increasing post-disaster network connectivity and its impact on biomass shipments has been generally ignored in the bio-energy supply chain literature. In light of this gap, this paper proposes a pre-disaster planning model that seeks to strengthen the multi-modal facility network links subject to budget constraints. The output of this model suggests optimal locations and capacities for multi-modal facilities and bio-refineries, the number of containers that can be efficiently transported among the facilities, and the optimal amount of biomass to be transported via trucks and multi-modal facilities under normal and disrupted scenarios. Additionally, the model identifies a subset of links that need to be strengthened to increase post-disaster connectivity even under a specified budget constraint. The aim is to make a right pre-disaster investment decision that reduces or eliminates risk and enhances post-disaster response, reconstruction, and recovery time for the bio-fuel supply chain network. The occurrences of disruption and/or failure of multi-modal facility links are characterized with a spatial statistics model. The resulting model can also accurately predict the failure probability of the facility links where historical data about disasters or disruptions may not be available.

Note that this proposed mixed-integer nonlinear programming (MINLP) model contains non-convexity term in the objective function and is considered a challenging problem as more of the solution space must be searched to ensure a global optimal solution. This motivates us to develop a generalized Benders decomposition algorithm that facilitates problem solving by using problem specific cuts, such as integer cuts, logistics cuts, and the

knapsack inequality. Multiple numerical experiments show that the accelerated generalized Benders decomposition algorithm is capable of producing a high quality feasible solution in a reasonable amount of time. In addition to proposing the model, another important contribution of this paper is its application to a real world case study. With data from Mississippi and Alabama, this study provides a number of new managerial insights that can effectively aid decision makers in designing a robust bio-fuel supply chain network. Suggestions about the optimal deployment of multi-modal facilities and bio-refineries and reliability improvement plans for transportation links are made, as well as what amounts of densified biomass and bio-fuel should be transported via different modes of transportation under both normal and disrupted scenarios. Finally, this paper examines the sensitivity of supply chain performance and multi-modal facility development to parameters, including link failure probabilities, reliability improvement costs, budget availability, irregularity in biomass supply, and the risk averseness degree of the decision maker.

The exposition of this paper is as follows. Section 2.2 formulates the mathematical model; Section 2.3 discusses the method of predicting the failure probability of links; Section 2.4 introduces the solution algorithm; Section 2.5 presents numerical results and draws managerial insights; and finally, Section 2.6 provides conclusions and future research directions.

2.2 Problem Description and Model Formulation

This paper builds a reliable model that aids the design and management of a bio-fuel supply chain network by considering the link failure probability among multi-modal fa-

cilities. The model addresses a pre-disaster planning problem that seeks to strengthen the links between the multi-modal facilities against disasters. The aim is to maximize post-disaster connectivity and minimize transportation costs between the origin, the location of the harvesting sites, to the destination, the markets where bio-fuel is sold. Figure 4.3 presents the bio-fuel logistic network consisting of biomass suppliers, demand cities and potential locations for bio-refineries and multi-modal facilities.

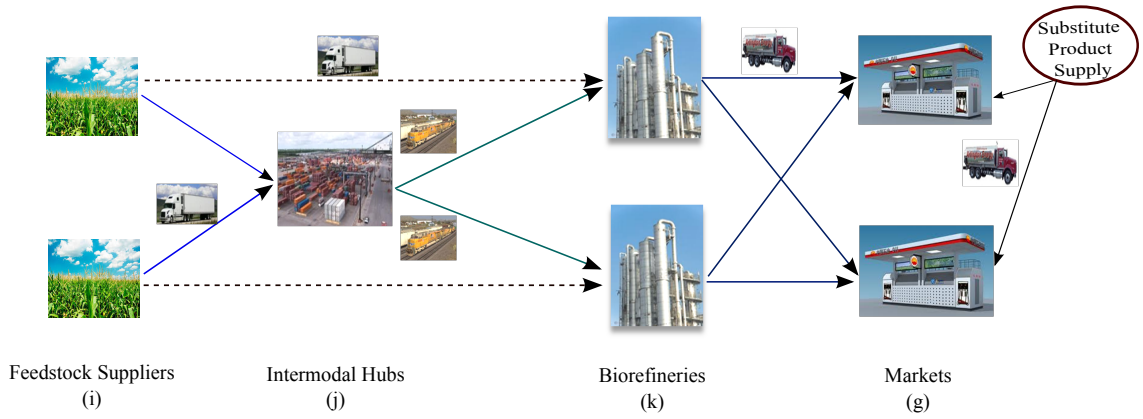


Figure 2.2

Network configuration of a bio-fuel supply chain

Consider a logistic network $G = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of arcs. Set $\mathcal{N}$ consists of the set of harvesting sites $\mathcal{I}$; the set of candidate multi-modal facility locations $\mathcal{J}$; the set of candidate biorefinery locations $\mathcal{K}$; and set of markets $\mathcal{G}$. Thus, the node set $\mathcal{N}$ is a union of all types of facilities in this logistics network, i.e., $\mathcal{N} = \mathcal{I} \bigcup \mathcal{J} \bigcup \mathcal{K} \bigcup \mathcal{G}$. Each site $i \in \mathcal{I}$ produces s_i unit of biomass. Each market $g \in \mathcal{G}$ demands b_g gallons of bio-fuel. This formulation assumes that a substitute product in the

market can be used to satisfy the demand for bio-fuel. The market price for the substitute product π_g represents the penalty cost per unit of unsatisfied demand at market $g \in \mathcal{G}$. This penalty also serves as a threshold of the unit biomass delivery cost, which indicates that if the unit biomass delivery cost exceeds this threshold, producing biofuel will be no longer profitable and thus demand will be satisfied by the substitute product. The notations used in this section are summarized in Table 2.2.

Locating an multi-modal facility/biorefinery of capacity level $l \in \mathcal{L}$ at each location $j \in \mathcal{J} \bigcup \mathcal{K}$ costs a fixed set-up cost of Ψ_{lj}. The set of arcs $\mathcal{A}$ is partitioned into four disjoint subsets, i.e., $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$, where $\mathcal{A}_1$ represents the set of arcs joining harvesting sites $\mathcal{I}$ with multi-modal facilities $\mathcal{J}$; $\mathcal{A}_2$ represents the set of arcs between multi-modal facilities $\mathcal{J}$ and biorefinery $\mathcal{K}$; $\mathcal{A}_3$ represents the set of arcs that directly connect from harvesting sites $\mathcal{I}$ to biorefinery $\mathcal{K}$; and finally, $\mathcal{A}_4$ represents the set of arcs that connect bio-refinery $\mathcal{K}$ with market $\mathcal{G}$. Transportation distance along arcs in $\mathcal{A}_1$ is short. Therefore, trucks are used to ship biomass along these arcs and its unit transportation cost is denoted by c_{ij}. On the other hand, transportation distances, as well as transportation quantities, along arcs in $\mathcal{A}_2$ are large. Thus, long haul transportation modes like rail can be used to ship biomass along these arcs. The unit cost along arcs $(j,k) \in \mathcal{A}_2$ are denoted by c_{jk}. Therefore, the unit transportation cost along arc $\{(i,j),(j,k)\}$ can be represented as $c_{ijk} = c_{ij} + c_{jk}$. Arcs in $\mathcal{A}_3$ represent direct shipments of biomass from suppliers to bio-fuel plants. These arcs are used when a supplier is located nearby a bio-refinery; thus, direct truck shipments are preferred. Finally, trucks are used to ship bio-fuel along arcs in $\mathcal{A}_4$. Densified biomass is usually transported in cargo containers between the multi-modal

facilities, and this process incurs a fixed loading and unloading cost denoted by ξ_{jk}. The capacity of each cargo container is denoted by v^{cap}.

Sets:

- $\mathcal{I}$: set of harvesting sites (farms)
- $\mathcal{J}$: set of multi-modal facilities
- $\mathcal{K}$: set of biorefineries
- $\mathcal{G}$: set of markets
- $\mathcal{L}$: set of capacities

Parameters:

- ψ_{lj} : fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$
- ψ_{lk} : fixed cost of opening a biorefinery of capacity $l \in \mathcal{L}$ at location $k \in \mathcal{K}$
- ξ_{jk} : fixed cost of a cargo container for transporting biomass along arc $(j,k) \in \mathcal{A}_2$
- c_{lk} : unit flow cost along arc $(l,k) \in \mathcal{A}$
- p_{lk} : unit production cost at a bio-fuel plant of size l
- s_i : amount of biomass available at site $i \in \mathcal{I}$
- b_g : bio-fuel demand of market $g \in \mathcal{G}$
- v^{cap} : cargo container capacity
- C_{lj} : biomass storage/handling capacity for a facility of size $l \in \mathcal{L}$ at location $j \in \mathcal{J} \bigcup \mathcal{K}$
- p_{lk}^{cap} : production capacity of a bio-fuel plant of size $l \in \mathcal{L}$ at location $k \in \mathcal{K}$
- B : total available fortification budget
- $\overline{\Delta}_{jk}$: maximum reliability that can be obtained with availability
- r_{jk} : cost associated with unit reduction of reliability in a given arc $(j,k) \in \mathcal{A}_2$
- π_g : unit penalty cost of not satisfying demand at market $g \in \mathcal{G}$
- ϕ : conversion rate from biomass to bio-fuel
- q_{jk} : failure probability associated with link $(j,k) \in \mathcal{A}_2$

Let q_{jk} is defined as the failure probability associated with link $(j,k) \in \mathcal{A}_2$. We assume that, when a link $(j,k) \in \mathcal{A}_2$ is disrupted, it is no longer available and the demand is fulfilled through emergency arcs $(i,k) \in \mathcal{A}_3$. For each link $(j,k) \in \mathcal{A}_2$, we introduce $\overline{\Delta}_{jk}$ as the maximum reliability improvement, which is obtainable with available technology. Therefore, ideally one can obtain up to $\Delta_{jk} = \overline{\Delta}_{jk}; \forall j \in \mathcal{J}, k \in \mathcal{K}$. We assume that the cost of a given link fortification is proportional to the reliability improvement of that link. We define r_{jk} as the cost associated with a unit reduction in the failure probability of link $(j,k) \in \mathcal{A}_2$. Let B denote the total available fortification budget. The following decision variables may now be introduced in this model.

The primary decision variables $\mathbb{Y} := \{Y_{lj}\}_{l \in \mathcal{L}, j \in \mathcal{J} \bigcup \mathcal{K}}$ determine the size and location for multi-modal facilities and biorefineries, i.e.,

$$Y_{lj} = \begin{cases} 1 & \text{if a multi-modal facility of size } l \text{ is used at location } j \\ 0 & \text{otherwise;} \end{cases}$$

$$Y_{lk} = \begin{cases} 1 & \text{if a biorefinery of size } l \text{ is opened at location } k \\ 0 & \text{otherwise;} \end{cases}$$

The second set of decision variables $\mathbb{Z} := \{Z_{jk}\}_{j \in \mathcal{J}, k \in \mathcal{K}}$ decides the number of containers to flow between each pair of facilities. The next set of decision variables determine how to route the biomass flowing from its origin to destination. Let $\mathbb{X} := \{X_{ijk}\}_{i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}} \cup \{X_{ik}\}_{i \in \mathcal{I}, k \in \mathcal{K}} \cup \{X_{kg}\}_{k \in \mathcal{K}, g \in \mathcal{G}}$ denote the flow of biomass through the multi-modal facilities, through highways, and the flow of bio-fuel from bio-refinery to the demand city. The final set of decision variables $\mathbb{P} := \{P_{lk}\}_{l \in \mathcal{L}, k \in \mathcal{K}}$ determine the amount of bio-fuel pro-

duced from a bio-finery k of capacity l, $\mathbb{U} := \{U_g\}_{g\in\mathcal{G}}$ determine the demand shortage of bio-fuel in market g, and $\mathbb{R} := \{\Delta_{jk}\}_{j\in\mathcal{J},k\in\mathcal{K}}$ determine the reliability improvement gains by fortifying arc $(j,k) \in \mathcal{A}_2$.

The problem is to identify the location and capacity of multi-modal facilities and bio-refineries among the candidate locations $j \in \mathcal{J} \bigcup \mathcal{K}$ and from the available capacities $l \in \mathcal{L}$. Also, the number of railcar used between the multi-modal facilities should be determined, as well as how much biomass to deliver through multi-modal facilities or through highways, how much bio-fuel to produce and ship from a bio-refinery to the demand city, and how much reliability improvement can be gained by fortifying the vulnerable links. The goal is to minimize the total system costs under normal and link failure scenarios. The following is the mixed integer nonlinear programming (MINLP) formulation for the reliable bio-fuel supply chain network system (**[RM]**):

$$\text{minimize}: \sum_{l\in\mathcal{L}} \sum_{j\in\mathcal{J}\bigcup\mathcal{K}} \psi_{lj} Y_{lj} + \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \xi_{jk} Z_{jk} + \sum_{i\in\mathcal{I}} \sum_{k\in\mathcal{K}} c_{ik} X_{ik} + \sum_{k\in\mathcal{K}} \sum_{g\in\mathcal{G}} c_{kg} X_{kg}$$

$$+ \sum_{g\in\mathcal{G}} \pi_g U_g + \sum_{l\in\mathcal{L}} \sum_{k\in\mathcal{K}} p_{lk} P_{lk} + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \Big(c_{ijk} \Big(1 - (q_{jk} - \Delta_{jk}) \Big) + c_{ik} \Big(q_{jk} - \Delta_{jk} \Big) \Big) X_{ijk}$$

Subject to

$$\sum_{k\in\mathcal{K}}\left(X_{ik}+\sum_{j\in\mathcal{J}}X_{ijk}\right) \leq s_i \qquad \forall i\in\mathcal{I} \tag{2.1}$$

$$\sum_{i\in\mathcal{I}}\phi\left(X_{ik}+\sum_{j\in\mathcal{J}}X_{ijk}\right) = \sum_{l\in\mathcal{L}}P_{lk} \qquad \forall k\in\mathcal{K} \tag{2.2}$$

$$\sum_{g\in\mathcal{G}}X_{kg} \leq \sum_{l\in\mathcal{L}}P_{lk} \qquad \forall k\in\mathcal{K} \tag{2.3}$$

$$\sum_{k\in\mathcal{K}}X_{kg}+U_g = b_g \qquad \forall g\in\mathcal{G} \tag{2.4}$$

$$\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijk} \leq \sum_{l\in\mathcal{L}}C_{lj}Y_{lj} \qquad \forall j\in\mathcal{J} \tag{2.5}$$

$$\sum_{i\in\mathcal{I}}X_{ik}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}X_{ijk} \leq \sum_{l\in\mathcal{L}}C_{lk}Y_{lk} \qquad \forall k\in\mathcal{K} \tag{2.6}$$

$$\sum_{i\in\mathcal{I}}X_{ijk} \leq v^{cap}Z_{jk} \qquad \forall j\in\mathcal{J},k\in\mathcal{K} \tag{2.7}$$

$$P_{lk} \leq p_{lk}^{cap}Y_{lk} \qquad \forall l\in\mathcal{L},k\in\mathcal{K} \tag{2.8}$$

$$\sum_{l\in\mathcal{L}}Y_{lj} \leq 1 \qquad \forall j\in\mathcal{J}\bigcup\mathcal{K} \tag{2.9}$$

$$\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}r_{jk}\Delta_{jk} \leq B \tag{2.10}$$

$$\Delta_{jk} \leq \overline{\Delta}_{jk} \qquad \forall j\in\mathcal{J},k\in\mathcal{K} \tag{2.11}$$

$$Y_{lj} \in \{0,1\} \qquad \forall l\in\mathcal{L},j\in\mathcal{J}\bigcup\mathcal{K} \tag{2.12}$$

$$Z_{jk} \in Z^{+} \qquad \forall j\in\mathcal{J},k\in\mathcal{K} \tag{2.13}$$

$$X_{ijk},X_{ik},X_{kg},P_{lk},U_g,\Delta_{jk} \geq 0 \qquad \forall i\in\mathcal{I},j\in\mathcal{J},k\in\mathcal{K} \tag{2.14}$$

In **[RM]**, the objective function minimizes the total expected system cost, including the expected construction and transportation costs under both normal and disruptive scenarios. More specifically, the first and second terms respectively represent the total set-up costs of establishing the multi-modal facilities and biorefineries and the fixed cost of transporting cargo containers between the multi-modal facilities. The third and fourth term represent

the cost associated with flow of biomass through direct route (i, k) and the flow of bio-fuel from bio-refinery k to demand city g. The fifth and sixth term represent the production cost of bio-fuel and penalty cost for demand shortage. The seventh term represents the regular transportation cost, which is weighted by $(1 - (q_{jk} - \Delta_{jk}))$, the probability that the link (j, k) is available. The final term of the objective function represents the transportation cost under disruption scenarios, where $(q_{jk} - \Delta_{jk})$ is the probability that the link (j, k) is unavailable.

Constraints (3.1) indicate that the amount of biomass shipped from a harvesting site $i \in \mathcal{I}$ is limited by its availability. Constraints (4.2) are the flow conservation constraints at bio-fuel plants. Constraints (4.3) indicate that the amount of bio-fuel delivered to the market is limited by the amount of bio-fuel produced at refinery $k \in \mathcal{K}$. Constraints (4.4) show that the total demand for bio-fuel will be fulfilled either through the distribution network or through substitute products available from the market. Constraints (4.5) demonstrate that the total amount of biomass shipped through multi-modal facility $j \in \mathcal{J}$ is limited by the facility capacity C_{lj}, $\forall l \in \mathcal{L}, j \in \mathcal{J}$. Similarly, constraints (4.6) indicate that the total amount of biomass shipped to a bio-refinery is limited by the refinery capacity C_{lk}, $\forall l \in \mathcal{L}, k \in \mathcal{K}$. Constraints (4.7) set a limit on the amount of biomass to be routed on the arcs $(j, k) \in \mathcal{A}_2$. Constraints (4.8) limit the bio-fuel production capacity at bio-refinery $k \in \mathcal{K}$. Constraints (4.9) indicate that, at most, one multi-modal facility/biorefinery of capacity $l \in \mathcal{L}$ is operating in a given location $j \in \mathcal{J} \bigcup \mathcal{K}$. Constraints (4.10) are the total fortification budget constraints. Constraints (4.11) set a limit on the amount of reliability that can be improved on the arcs $(j, k) \in \mathcal{A}_2$. Finally, constraints (4.12) are the binary

constraints; (4.13) are the integrality constraints; and (3.14) are the standard non-negativity constraints. Additionally, in order to improve the performance of the branch-and-bound process we add the following valid inequalities in model **[RM]**.

$$Z_{jk} \leq \sum_{l \in \mathcal{L}} \lceil \frac{C_{lj}}{v_{jk}^{cap}} \rceil Y_{lj} \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.15}$$

$$Z_{jk} \leq \sum_{l \in \mathcal{L}} \lceil \frac{C_{lk}}{v_{jk}^{cap}} \rceil Y_{lk} \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.16}$$

Constraints (4.18) and (4.19) indicate that if no hub is established at location j and k, then no containers will flow between arc (j, k). The constraints also impose the maximum number of containers that can be transported along an active arc.

2.3 Characterizing Link Failure Probability q_{jk}

We estimate the link failure probability q_{jk} based on a database that consists of historical events of disasters. We assume that whenever a disaster occurs, it causes disruption/failure of infrastructure (including links and facilities) within the affected areas. In other words, a link's disruption probability is approximately equivalent to the probability of a disaster.

Previous work [76] estimates the failure probability of facilities by assuming that disaster events at different locations are independent, and the likelihood of the occurrence of a disaster is approximated using the relative frequency. This approach may be valid when the facilities are distant from each other. However, due to the proximity of links and facilities and the probability of disaster, the probability of link failure probability may not be spatially independent. For example, when a disaster, flood for instance, occurs in a

specific area, the infrastructure close to this area is more likely to be affected/disrupted. On the other hand, the probability that distant infrastructure is affected by this flood is much smaller. Therefore, characterizing the spatial dependency among the probability of the occurrence of disasters, and thus the disruption/failure probability of the infrastructure, is crucial to predicting the failure probability of infrastructure, where the historical disaster/disruption data is unavailable.

This paper generalizes the failure probability estimation of facilities in a previous work to those of links by accounting for the spatial dependence among the failure probabilities at different locations. To do this, each link is divided into multiple sections, and the failure probability of each section is estimated. Recall that each link, connecting two facilities, may cover long distance, e.g., 200 miles. The failure probability at a specific location of link may not well characterize the failure probability of the entire link since failure probability at different locations may vary. To address this challenge, each link is segmented into multiple exclusive sections, and the failure probability at each section of the link is considered. The maximum length of a section, denoted by d_a, is selected in a way that the failure probability does not vary much within each segment. In other words, d_a represents the diameter of the *Disaster Affected Zone* (DAZ), and the choice of d_a depends on the major type of disasters that disrupt biomass infrastructure. For instance, the value of d_a chosen for the state of Mississippi is 30 miles because most infrastructure disruptions are caused by flood and tornado. When such disasters damage certain locations, the average diameter of the affected zone is approximately 30 miles [52].

Once we estimate the failure probability of the sections within each link, the resulting probability is used to assess the probability of link failure. This paper, utilizes a conservative estimate for link failure probability by evaluating it using the maximum failure probability of sections within each link. Assume that link (j, k) can be divided into n_{jk} sections, each of which has a failure probability of q_ℓ for $\ell = 1, \ldots, n_{jk}$. Thus, link failure probability q_{jk} can be estimated as $q_{jk} = \max_{\ell=1,\ldots,n_{jk}} q_\ell$.

In what follows, we utilize a spatial statistic model to estimate the failure probability of segments for each link q_ℓ by leveraging the historical data of disaster occurrence. In a previous work [76], the relative frequency of disasters at a specific location has been used to evaluate the risk of disaster which reveals the failure probability of infrastructure at a specific location. This work utilizes a spatial statistics model to interpolate the failure probability of locations, where historical data is unavailable.

2.3.1 Spatial Model of Link Failure Probability

We propose a continuous spatial statistics model to characterize the spatial correlation and trend of failure probability. The model is described below:

$$q(\mathbf{s}) \quad = \quad \mu + e(\mathbf{s}) \tag{2.17}$$

where,

- **s** represents the long-lat coordinates for the possible location of biomass infrastructure in the area of interest. Specifically, **s** can represent the long-lat coordinates of a facility, a segment of facility-to-facility link, highway, etc.

- μ, an unknown parameter to be estimated from the data, denotes the average probability of disaster occurrence and thus the average of failure probability of biomass infrastructure caused by the disruption of disasters. μ is an unknown parameter to be estimated from the data.

- $e(\mathbf{s})$ captures the variability in the failure probability and accounts for the spatial dependence in failure probability $q(\mathbf{s})$. In general, the correlation between the failure probability of locations with close physical proximity tends to be stronger than those distant to each other.

We assume that $e(\mathbf{s})$ is a spatial process with a mean of 0 and a covariance structure of Σ. Σ is unknown in most cases and can be estimated based on a historical data set. We characterize the structure dependence of Σ via the semi-variogram function $\gamma(\mathbf{h})$ defined below:

$$\gamma(\mathbf{h}) \quad = \quad \frac{1}{2}\mathrm{Var}\left[q(\mathbf{s}) - q(\mathbf{s}+\mathbf{h})\right] \tag{2.18}$$

Given the observed historical data of failure probability, $q(\mathbf{s}_1), \ldots, q(\mathbf{s}_n)$. The Matheron empirical estimate of $\gamma(\mathbf{h})$ is

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|}\sum_{N(\mathbf{h})}\{q(\mathbf{s}_i) - q(\mathbf{s}_j)\}^2 \tag{2.19}$$

where $N(\mathbf{h})$ is the set of location pairs $(\mathbf{s}_i, \mathbf{s}_j)$ with a coordinate difference $\mathbf{h}$, and $|N(\mathbf{h})|$ is the number of distinct pairs in this set.

2.3.2 Predicting The Link Failure Probability

The model of $\gamma(\mathbf{h})$ can be established based on its empirical version $\hat{\gamma}(\mathbf{h})$ by utilizing various methods of model fitting such as Maximum Likelihood Estimate, Least Square Estimate, Weighted Least Square Estimate, and others [21]. Once the semi-variogram model of γ is established, the resulting model, by means of Kriging, is used to interpolate the failure probability at given locations where historical data is unavailable. Since the mean value μ is unknown, the Ordinary Kriging can be used. Let $\mathbf{q} = (q(\mathbf{s}_1), \ldots, q(\mathbf{s}_n))'$ be the set of available historical data set of failure probability. The Ordinary Kriging estimator for a location $\mathbf{s}_0$ is a linear predictor in the form of $\hat{q}_{OK}(\mathbf{s}_0) = \lambda'\mathbf{q}$, where λ denotes the unknown weight coefficients to be estimated. To this end, we choose weight coefficients $\lambda = (\lambda_1, \ldots, \lambda_n)'$ that minimize the loss function shown below:

$$L = \mathbb{E}[(\lambda' q - q(s_0))^2]$$

subject to the weight constraint $\sum_{i=1}^{n} \lambda_i = 1$.

This problem can be solved as an unconstrained minimization problem introducing the Lagrange multiplier θ:

$$\arg\min_{\lambda,\theta} L \quad = \quad \arg\min_{\lambda,\theta} \left\{ \mathbb{E}[(\lambda' q - q(\mathbf{s}_0))^2] - 2\theta \left(\sum_{i=1}^{n} \lambda_i - 1 \right) \right\} \tag{2.20}$$

Let $\Gamma = [\gamma(\mathbf{s}_i - \mathbf{s}_j)]$ be the $n \times n$ semi-variogram matrix of the existing data points and $\gamma(\mathbf{s}_0) = [\gamma(\mathbf{s}_0 - \mathbf{s}_1), \ldots, \gamma(\mathbf{s}_0 - \mathbf{s}_n)]'$ be the semi-variogram vector between $\mathbf{s}_0$ and $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. The resulting Ordinary Kriging estimate, expressed in terms of the semi-variogram γ, is:

$$\hat{q}_{OK}(\mathbf{s}_0) = \lambda' q(\mathbf{s}) \tag{2.21}$$

where

$$\lambda' = \left(\gamma(\mathbf{s}_0) + 1\frac{1 - 1'\Gamma^{-1}\gamma(\mathbf{s}_0)}{1'\Gamma^{-1}1}\right)' \Gamma^{-1} \tag{2.22}$$

In real world applications, the true form of the semi-variogram function γ is usually unknown. In this case, the empirical model of γ, namely $\hat{\gamma}$ as shown in expression (2.19), can be used in expression (2.21).

2.4 Solution Algorithm

2.4.1 Generalized Benders Decomposition Algorithm

Due to the presence of non-convexity term in the objective function of model **[RM]**, solving this problem is computationally expensive as more of the solution space must be searched to ensure a global optimal solution. Therefore, we propose a Generalized Benders decomposition algorithm [40] to solve this challenging mixed integer non-linear programming (MINLP) model **[RM]**. This approach works well under the assumption that when the complicated variables are temporarily fixed, the restricted problem is easy to solve. The aim is to provide a high quality feasible solution for model **[RM]** in a reasonable amount of time. For fixed values of $\{\hat{Y}_{lj}\}_{l\in\mathcal{L},j\in\mathcal{J}\bigcup\mathcal{K}}$ and $\{\hat{Z}_{jk}\}_{j\in\mathcal{J},k\in\mathcal{K}}$ the Benders subproblem **[BD(SUB)]** can now be written as follows:

[BD(SUB)] $Minimize \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} c_{ik}X_{ik} + \sum_{k\in\mathcal{K}}\sum_{g\in\mathcal{G}} c_{kg}X_{kg} + \sum_{g\in\mathcal{G}} \pi_g U_g + \sum_{l\in\mathcal{L}}\sum_{k\in\mathcal{K}} p_{lk}P_{lk} +$

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\Bigg(c_{ijk}\Big(1-(q_{jk}-\Delta_{jk})\Big) + c_{ik}\Big(q_{jk}-\Delta_{jk}\Big)\Bigg)X_{ijk}$$

Subject to (3.1)-(4.4), (4.10), (4.11), (3.14), and

$$\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijk} \leq \sum_{l\in\mathcal{L}} C_{lj}\hat{Y}_{lj} \qquad \forall j\in\mathcal{J} \tag{2.23}$$

$$\sum_{i\in\mathcal{I}} X_{ik} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} X_{ijk} \leq \sum_{l\in\mathcal{L}} C_{lk}\hat{Y}_{lk} \qquad \forall k\in\mathcal{K} \tag{2.24}$$

$$P_{lk} \leq p_{lk}^{cap}\hat{Y}_{lk} \qquad \forall l\in\mathcal{L}, k\in\mathcal{K} \tag{2.25}$$

$$\sum_{i\in\mathcal{I}} X_{ijk} \leq v^{cap}\hat{Z}_{jk} \qquad \forall j\in\mathcal{J}, k\in\mathcal{K} \tag{2.26}$$

$$Y_{lj} = Y_{lj}^n \ : \ \delta_{lj} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J} \tag{2.27}$$

$$Y_{lk} = Y_{lk}^n \ : \ \lambda_{lk} \qquad \forall l\in\mathcal{L}, k\in\mathcal{K} \tag{2.28}$$

$$Z_{jk} = Z_{jk}^n \ : \ \gamma_{jk} \qquad \forall j\in\mathcal{J}, k\in\mathcal{K} \tag{2.29}$$

For a given value of $\{\hat{Y}_{lj}\}_{l\in\mathcal{L}, j\in\mathcal{J}\bigcup\mathcal{K}}$ and $\{\hat{Z}_{jk}\}_{j\in\mathcal{J},k\in\mathcal{K}}$ problem **[BD(SUB)]** is always feasible. This is ensured via constraints (4.4). Let $\{\delta_{lj}\geq 0 | l\in\mathcal{L}, j\in\mathcal{J}\}$, $\{\lambda_{lk}\geq 0 | l\in\mathcal{L}, k\in\mathcal{K}\}$, and $\{\gamma_{jk}\geq 0 | j\in\mathcal{J}, k\in\mathcal{K}\}$ be the dual variables associated with constraints (4.23)-(4.25), respectively. These variables are required to construct Benders optimality cut (shown below via constraints (4.26)). Introducing an additional variable θ, the underlying Benders *master problem* **[BD(M)]** for the generalized Benders decomposition algorithm can now be written as follows:

[BD(M)] $Minimize \sum_{l\in\mathcal{L}} \sum_{j\in\mathcal{J}\bigcup\mathcal{K}} \psi_{lj} Y_{lj} + \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \xi_{jk} Z_{jk} + \theta$

Subject to (4.9), (4.12), (4.13), (4.18), (4.19), and

$$\begin{aligned} \theta &\geq v(\mathbf{Y}^n, \mathbf{Z}^n) + \sum_{l\in\mathcal{L}} \sum_{j\in\mathcal{J}} \delta^n_{lj}(Y_{lj} - Y^n_{lj}) + \sum_{l\in\mathcal{L}} \sum_{k\in\mathcal{K}} \lambda^n_{lk}(Y_{lk} - Y^n_{lk}) \\ &+ \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \gamma^n_{jk}(Z_{jk} - Z^n_{jk}) \qquad \forall n \in N \end{aligned} \tag{2.30}$$

$$\theta \geq 0 \tag{2.31}$$

where $\{Y^n_{lj}\}_{l\in\mathcal{L}, j\in\mathcal{J}\bigcup\mathcal{K}}$ and $\{Z^n_{jk}\}_{j\in\mathcal{J}, k\in\mathcal{K}}$ are the y-vector and z-vector generated at iteration $n \in N$ and N denotes the number of Benders cuts generated so far. Constraints (4.26) are referred to as optimality cut constraints which can be generated iteratively using the optimal objective function value $v(\mathbf{Y}^n, \mathbf{Z}^n)$ obtained from solving the Benders subproblem **[BD(SUB)]** and the dual vectors of δ, λ, and γ. The overall algorithm is described as follows:

The solution (z^n_{MP}) obtained from the master problem **[BD(M)]** provides a lower bound (LB^n) for the generalized Benders decomposition algorithm. At each iteration, subproblem **[BD(SUB)]** is evaluated by fixing the values of $\{Y^n_{lj}\}_{l\in\mathcal{L}, j\in\mathcal{J}\bigcup\mathcal{K}}$ and $\{Z^n_{jk}\}_{j\in\mathcal{J}, k\in\mathcal{K}}$ obtained from the master problem **[BD(M)]**.

Let $z^n_{MAS} = \sum_{l\in\mathcal{L}} \sum_{j\in\mathcal{J}\bigcup\mathcal{K}} \psi_{lj} Y_{lj} + \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \xi_{jk} Z_{jk}$ and $v(\mathbf{Y}^n, \mathbf{Z}^n)$ denote the solution of subproblem **[BD(SUB)]** in iteration n of the Benders decomposition algorithm. In each iteration, the summation of z^n_{MAS} and $v(\mathbf{Y}^n, \mathbf{Z}^n)$ provides an upper bound (UB^n) for model **[RM]**. At the end of each iteration we check if the difference between the upper and lower bound falls below a threshold limit ϵ. Otherwise, constraints (4.26) are added to

the master problem **[BD(M)]**. A pseudo-code of the generalized Benders decomposition algorithm is provided in **Algorithm 1**.

Algorithm 1: Generalized Benders Decomposition

$UB^n \leftarrow +\infty$, $LB^n \leftarrow -\infty$, $n \leftarrow 1$, ϵ, *terminate* $\leftarrow$ **false**
while (*terminate* $=$ **false**) **do**
  Solve **[BD(M)]** to obtain $\{Y_{lj}^n\}_{l\in\mathcal{L},j\in\mathcal{J}\bigcup\mathcal{K}}$, $\{Z_{jk}^n\}_{j\in\mathcal{J},k\in\mathcal{K}}$, z_{MP}^n, and z_{MAS}^n
  if ($z_{MP}^n > LB^n$) **then**
    $LB^n \leftarrow z_{MP}^n$
  end if
  Solve sub-problem **[BD(SUB)]** to obtain $v(\mathbf{Y}^n, \mathbf{Z}^n)$
  Calculate δ_{lj}^n, λ_{lk}^n, and γ_{jk}^n using constraints (4.23)-(4.25)
  if (z_{MAS}^n + $v(\mathbf{Y}^n, \mathbf{Z}^n) < UB^n$) **then**
    $UB^n \leftarrow z_{MAS}^n + v(\mathbf{Y}^n, \mathbf{Z}^n)$
  end if
  if ($(UB^n - LB^n)/UB^n \leq \epsilon$) **then**
    terminate $\leftarrow$ **true**
  else
    Add cut (4.26) in **[BD(M)]**
  end if
  $n \leftarrow n + 1$
end while

2.4.2 Algorithmic Enhancement

In this section, we study how to improve the computational performance of the generalized Benders decomposition algorithm in solving model **[RM]**. To achieve this goal, we have developed few enhancement strategies to improve the performance of the generalized Benders decomposition algorithm.

- **Integer cuts:** To expedite the running time of the master problem we add the following inequalities generated using a local branching technique [35]. These inequalities excludes all binary values searched in the previous iterations and force the master

problem to select a solution different from the previous solutions. To add these inequalities lets first define $\{\hat{Y}^n_{lj}\}_{\forall l \in \mathcal{L}, j \in \mathcal{J} \bigcup \mathcal{K}}$ be the solutions obtained from solving the master problem in iteration n. Let $\mathcal{Y}^n_1 = \{(l,j) | \hat{Y}^n_{lj} = 1, \forall l \in \mathcal{L}, j \in \mathcal{J}\}$ and $\mathcal{Y}^n_2 = \{(l,k) | \hat{Y}^n_{lk} = 1, \forall l \in \mathcal{L}, k \in \mathcal{K}\}$. We add the following constraint to the master problem in iteration $n+1$:

$$\sum_{(l,j) \in \mathcal{Y}^n_1} (1 - Y_{lj}) + \sum_{(l,k) \in \mathcal{Y}^n_2} (1 - Y_{lk}) + \sum_{(l,j) \notin \mathcal{Y}^n_1} Y_{lj} + \sum_{(l,k) \notin \mathcal{Y}^n_2} Y_{lk} \quad \geq \quad 1 \quad (2.32)$$

These inequalities when added to the master problem during iteration $n+1$, force the problem to generate a new solution different from the solution generated during iteration n.

- **Logistics Constraints:** In the initial stages of the Benders decomposition algorithm, the master problem produces very few first-stage decision variables. This process continues until sufficient information is gathered from the subproblem via constraints (4.26). To overcome this issue in each iteration of the Benders decomposition algorithm we add the following logistics constraints in the master problem. Doing this will improve the running time of the Benders decomposition algorithm.

  Recall the demand satisfying constraints described in equation (4.4). Since this is a minimization problem, these constraints can be expressed as follows:

$$\sum_{k \in \mathcal{K}} X_{kg} + U_g \quad \geq \quad b_g \qquad\qquad \forall g \in \mathcal{G} \qquad (2.33)$$

  Furthermore, constraints (4.3) and (4.8) can be rewritten as follows:

$$\sum_{g \in \mathcal{G}} X_{kg} \quad \leq \quad \sum_{l \in \mathcal{L}} p^{cap}_{lk} Y_{lk} \qquad\qquad \forall k \in \mathcal{K} \qquad (2.34)$$

Combining constraints (4.28) and (4.29) and dropping the penalty term U_g obtains the following valid-inequalities that are added to the master problem:

$$\sum_{l \in \mathcal{L}, k \in \mathcal{K}} p_{lk}^{cap} Y_{lk} \;\; \geq \;\; \sum_{g \in \mathcal{G}} \bar{b}_g \tag{2.35}$$

Here $\bar{b}_g$ represents the amount of bio-fuel demand expected to be met through the supply chain network. Thus, $\bar{b}_g = \alpha * b_g$, $\forall g \in \mathcal{G}$. We can initialize the value of α between 0.0 to 1.0; and when $\alpha = 1.0$, constraints (4.30) require that all the demand is satisfied by this supply chain and the introduction of variable U_g becomes redundant. In our application, we initially set $\alpha = 0.2$ to provide the Benders master problem few initial location decisions; however, we remove this constraints when the gap between the upper and lower bound of the overall Benders decomposition algorithm falls below 20%.

- **Knapsack Inequalities:** In order to speed up the branch-and-bound procedure in the solver, we have added the following knapsack inequality in the master problem **[BD(M)]** of the generalized Benders decomposition algorithm. These cuts will help the state-of-the-art solvers such as CPLEX to derive a variety of valid inequalities which will expedite the convergence of the Benders decomposition algorithm. Therefore, we add the following cuts to the master problem **[BD(M)]** in iteration $n + 1$:

$$LB^n \;\; \leq \;\; \sum_{l \in \mathcal{L}} \sum_{j \in \mathcal{J} \bigcup \mathcal{K}} \psi_{lj} Y_{lj} + \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \xi_{jk} Z_{jk} + \theta \tag{2.36}$$

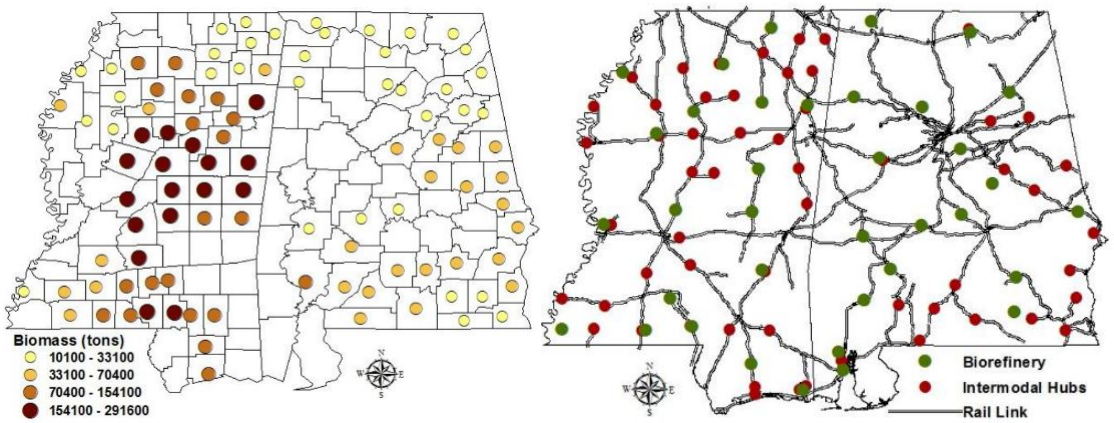
where LB^n denotes the best known lower bound obtained so far.

- **Obtaining Good Solutions before Convergence:** In the initial stage of the Benders decomposition algorithm, the master problem typically produces low quality solutions. The process continues until sufficient information from the subproblem is passed to the master problem via constraints (4.26). Additionally, the master problem is an integer problem for which generating an optimal solution for even a moderate size network problem is a challenging task. In order to alleviate this problem, we initially set a large optimality gap which we gradually reduced as the algorithm progressed. For instance, initially we set an optimality gap of 5% which is reduced to 1% when the gap between the upper and lower bound of the Benders decomposition algorithm falls below 10%.
- **Setting Branching Priorities:** In order to accelerate the solution of the master problem, we set proper branching priorities for variables Z_{jk}, Y_{lk} and Y_{lj}. Branching priorities helps CPLEX find the order in which the solver branches these variables. Numerical analysis indicates that setting branching on Z_{jk} variables first followed by Y_{lk} and Y_{lj} saves some computational time when solving the master problem of the Benders decomposition algorithm.

2.5 Computational Study and Managerial Insights

This section summarizes and interprets the results from the numerical study. We use the states of Mississippi and Alabama as a testing ground for our model. This case study demonstrates how the model can be applied to relevant real-world network design problems. The proposed model and algorithm are programmed in GAMS 24.2.1 [39] and ex-

ecuted on a desktop computer with Intel Core i7 3.50 GHz processor and 32.0 GB RAM. The optimization solvers used are ILOG CPLEX 12.6, BONMIN [9], and CONOPT 3.15 [37].



(a) Biomass distribution (b) Network representation

Figure 2.3

Biomass distribution and facility locations

2.5.1 Data Description

***Biomass Supply and Demand*:** The region of interest in this case study consists of the states of Mississippi and Alabama. The two main biomass feedstocks in this region are corn stover and forest residues. The biomass availability data is provided by the Knowledge Discovery Framework (KDF) database of United States Department of Energy [10]. This data was further processed by Idaho National Lab (INL) to identify the amount of densified biomass available in this region. Figure 4.5a shows the distribution of densified biomass available in this region. Every year, this region produces 6.85 million tons (MT) of

densified biomass from 99 different counties. Note that only those counties that produces more than 10,000 tons of biomass each year are considered here.

The total annual bio-fuel demand for our test region is set at 800 million gallon per year (MGY) [113]. Counties with a population greater than 30,000 are considered the demand points in this study. Based on this criteria, 41 counties from the region are selected. Assume that the distribution of the population in a particular region is a good indicator of the distribution of demand for bio-fuel. We use the centroid of the county as the point to which bio-fuel is to be delivered.

***Investment Costs*:** We consider a total of 71 potential multi-modal facility locations (rail ramps) and 36 potential bio-refinery locations. Figure 4.5b presents the potential locations of these multi-modal facilities and bio-refineries. We consider five different bio-refinery sizes: 20 MGY, 40 MGY, 60 MGY, 100 MGY, and 150 MGY and five different rail ramp capacities: 0.6 MTY, 0.8 MTY, 0.9 MTY, 1.05 MTY, and 1.20 MTY. The annualized fixed costs for facilities and bio-refineries with their available capacities are obtained from the study of Marufuzzaman et al. [76]. Although the actual fixed cost may vary by location, a common fixed cost is presented here as a reasonable approximation.

***Transportation Costs*:** This study assumes that trucks are used to transport biomass from a feedstock supplier $i \in \mathcal{I}$ to a multi-modal facility $j \in \mathcal{J}$. Trucks can also be used to transport biomass directly from a feedstock supplier $i \in \mathcal{I}$ to a bio-refinery $k \in \mathcal{K}$ and bio-fuel from bio-refinery $k \in \mathcal{K}$ to market $g \in \mathcal{G}$. The major cost components needed to calculate the unit cost of truck transportation are summarized in Table 4.2.

Table 2.1

Truck transportation cost components [88]

	Feedstock		Bio-fuel	
Item	Value	Unit	Value	Unit
Loading/unloading	5.0	$/wet ton	0.02	$/gallon
Time dependent	29.0	$/hr/truckload	32.0	$/hr/truckload
Distance dependent	1.20	$/mile/truckload	1.3	$/mile/truckload
Truck capacity	25	wet tons/truckload	8,000	gallons/truckload
Average travel speed	40	miles/hour	40	miles/hour

We set the fixed shipment cost per rail car equal to $2,248 ($\xi_{jk}$), and the unit transportation cost per rail car per mile is equal to $1.12. These cost parameters are based on a single rail car of capacity 100 tons and are obtained from the study of Gonzales et al. [42]. We used Arc GIS Desktop 10 to create a transportation network and then used this network to identify the shortest paths. The network includes existing railways; local, rural, and urban roads along with the major highways in Mississippi and Alabama.

***Estimating Disruption Probabilities*:** Failure probabilities of multi-modal facilities are obtained from a study by Marufuzzaman et al. [76]. In developing these estimates, the authors consider three major types of natural disasters (hurricanes, floods and droughts) that affect the Southeast U.S. Figure 2.4 presents the estimated disruption probabilities of the candidate multi-modal facilities where the size of a circle is proportional to its disruption probability. From the figure it is observed that the multi-modal facilities located in the coastal region of Mississippi and Alabama face high disruption risk.

To model the spatial distribution of failure probability and predict it in locations where historical data is unavailable, we first estimate the semi-variogram of failure probability $q(s)$ based on expression (2.19). The resulting empirical semi-variogram is demonstrated

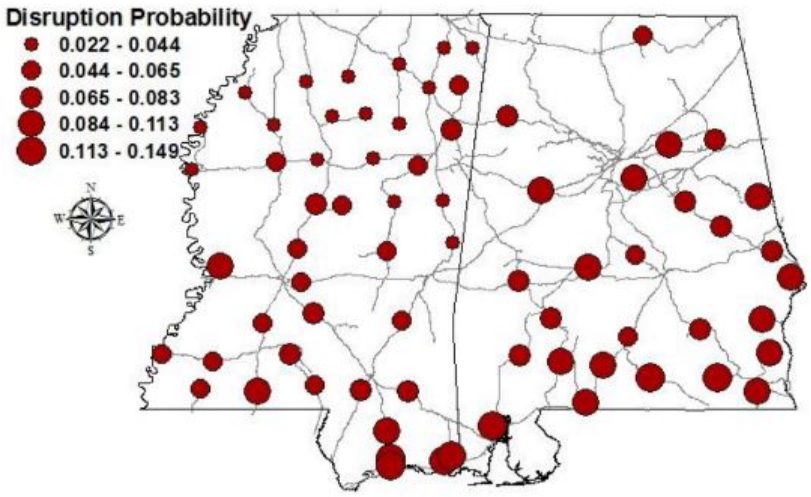


Figure 2.4

Disruption probability of multi-modal facilities [76]

in Figure 2.5. For both plots, the vertical axis represents the calculated $\hat{\gamma}$ and the horizontal axis represents the distance between any pair of locations (s_i, s_j). The plot on the left is the scatter plot which is calculated $\hat{\gamma}$ based on all pairs of (s_i, s_j)'s, whereas, the plot on the right is the binned version of the scatter plot. Based on its plot, we fit $\hat{\gamma}$ using the method of Weighted Least Square estimate. The fitted model of $\hat{\gamma}$ is shown in Figure 2.6, which shows that a linear model may be a good fit for the semi-variogram function based on the existing data set. We also examine the fitted model using the leave-one-out cross-validation strategy. To do this, data points are removed one by one and predicted by Kriging using the remaining data. Graphical results of the leave-one-out cross-validation are shown in Figure 2.7, which summarizes the residual error of leave-one-out cross-validation. It is shown in Figure 2.7 that most of residual prediction errors are around the value of 0, which is supportive of the estimated semi-variogram model. Once the semi-variogram

model is established, we use the Ordinary Kriging method, as shown in Equation (2.21), to interpolate failure probability of the links.

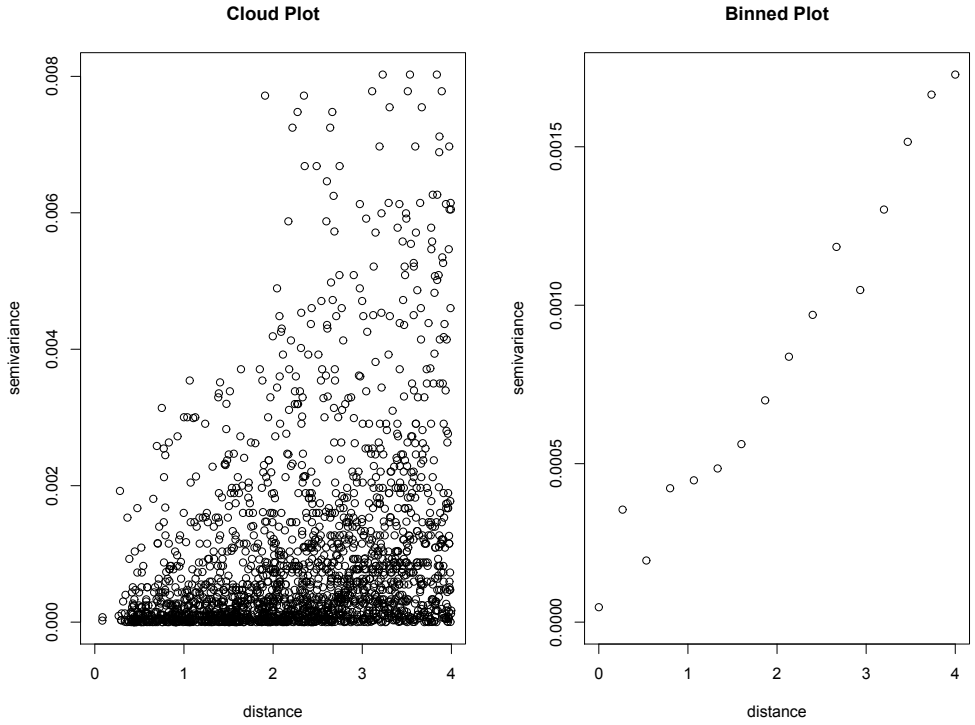


Figure 2.5

Empirical plot of the semi-variogram $\hat{\gamma}$

2.5.2 Results and Discussion

Impact of budget allocation on system performance: Figure 2.8 depicts the bio-fuel supply chain network under various budget availability scenarios: $5 million and $10 million. These prospective budgets total the amount of money given to the decision maker to invest

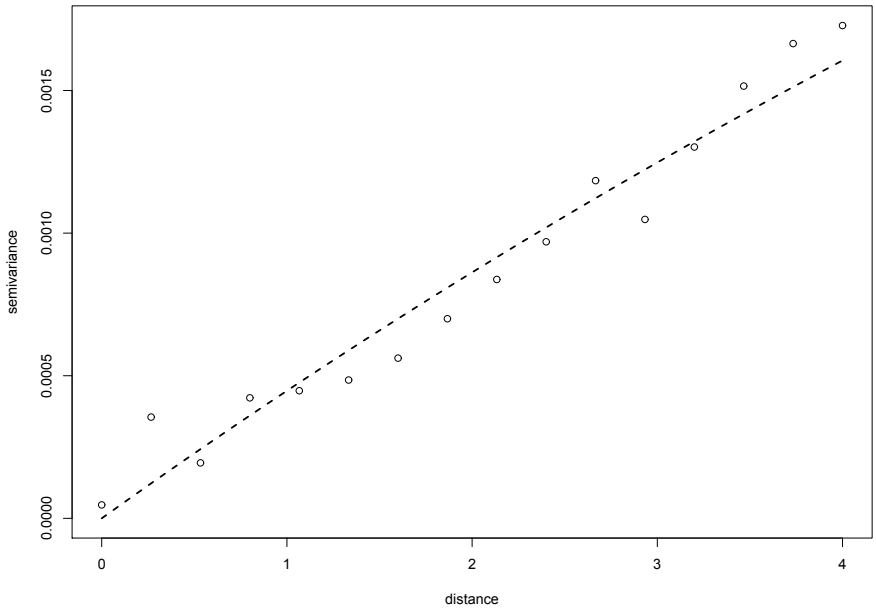


Figure 2.6

Fitting the semi-variogram model

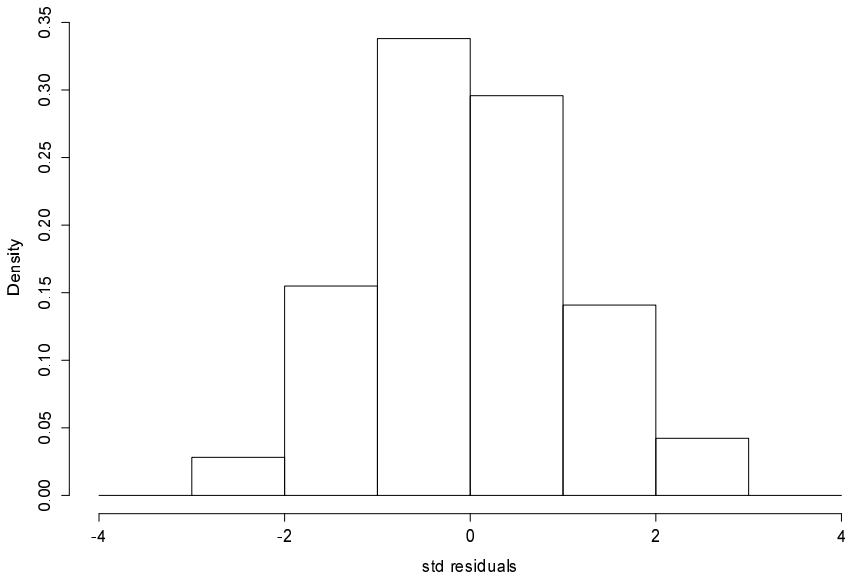
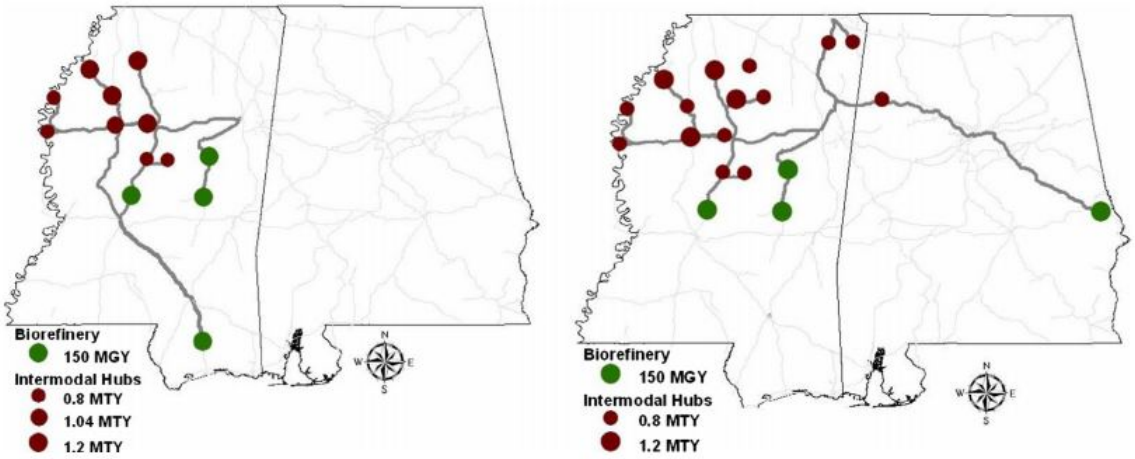


Figure 2.7

Leave-one-out cross validation for the semi-variogram model



(a) Budget = $5 million (b) Budget = $10 million

Figure 2.8

Network representation under various budget availability scenarios

in improving the reliability of multi-modal facility-to-facility links. The figure shows that budget availability has a significant impact on the location and number of facilities and the reliability of link improvement plans. For example, model **[RM]** selects nine multi-modal facilities with a $5 million budget, whereas the model selects fifteen multi-modal facilities with $10 million. Additionally, the likely reliability improvement suggested by the higher budget scenario is greater than that of lower budget scenario on the same links. With the availability of higher budgets, the model tends to cover more regions as the reliability of the links between the facilities improves. Note that model **[RM]** selects most of the multi-modal facilities in the state of Mississippi because 82.06% of the total availability of biomass in the two states is located in Mississippi.

Impact of reliability improvement cost (r_{jk}) on system performance: Figure 2.9 shows how reliability improvement cost impacts unit bio-fuel cost, the number of facilities operating,

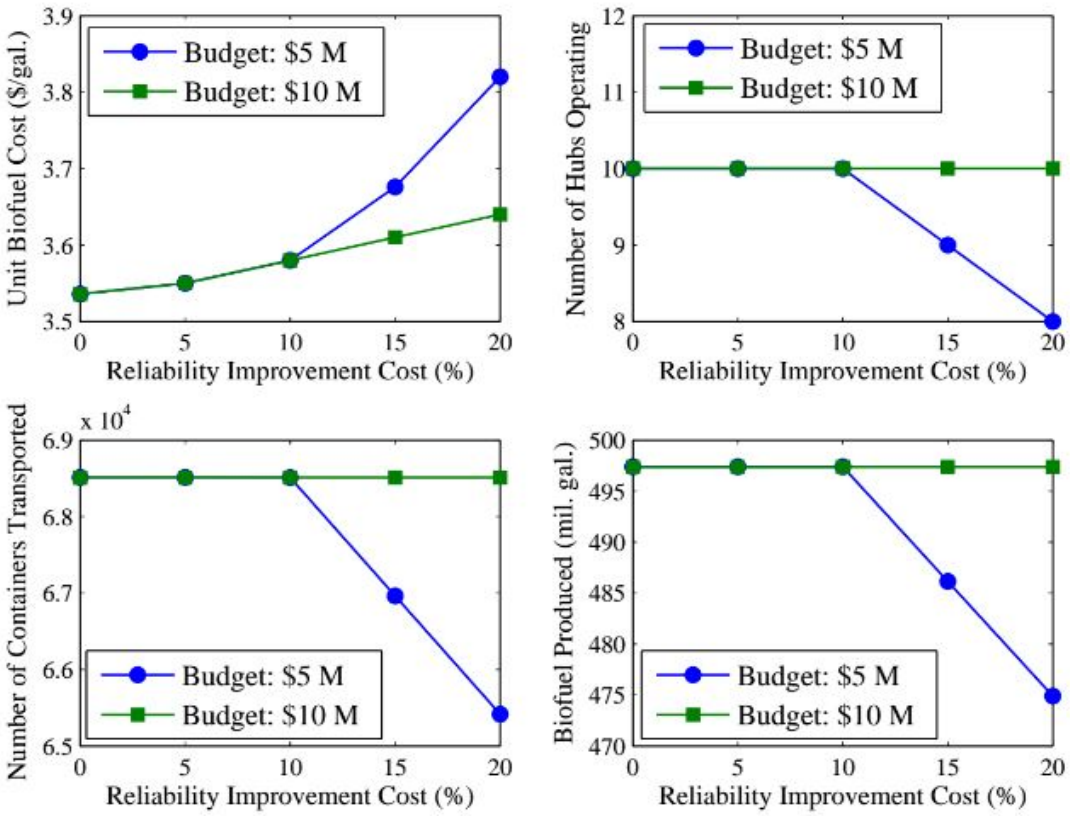


Figure 2.9

Impact of reliability improvement cost on system performance

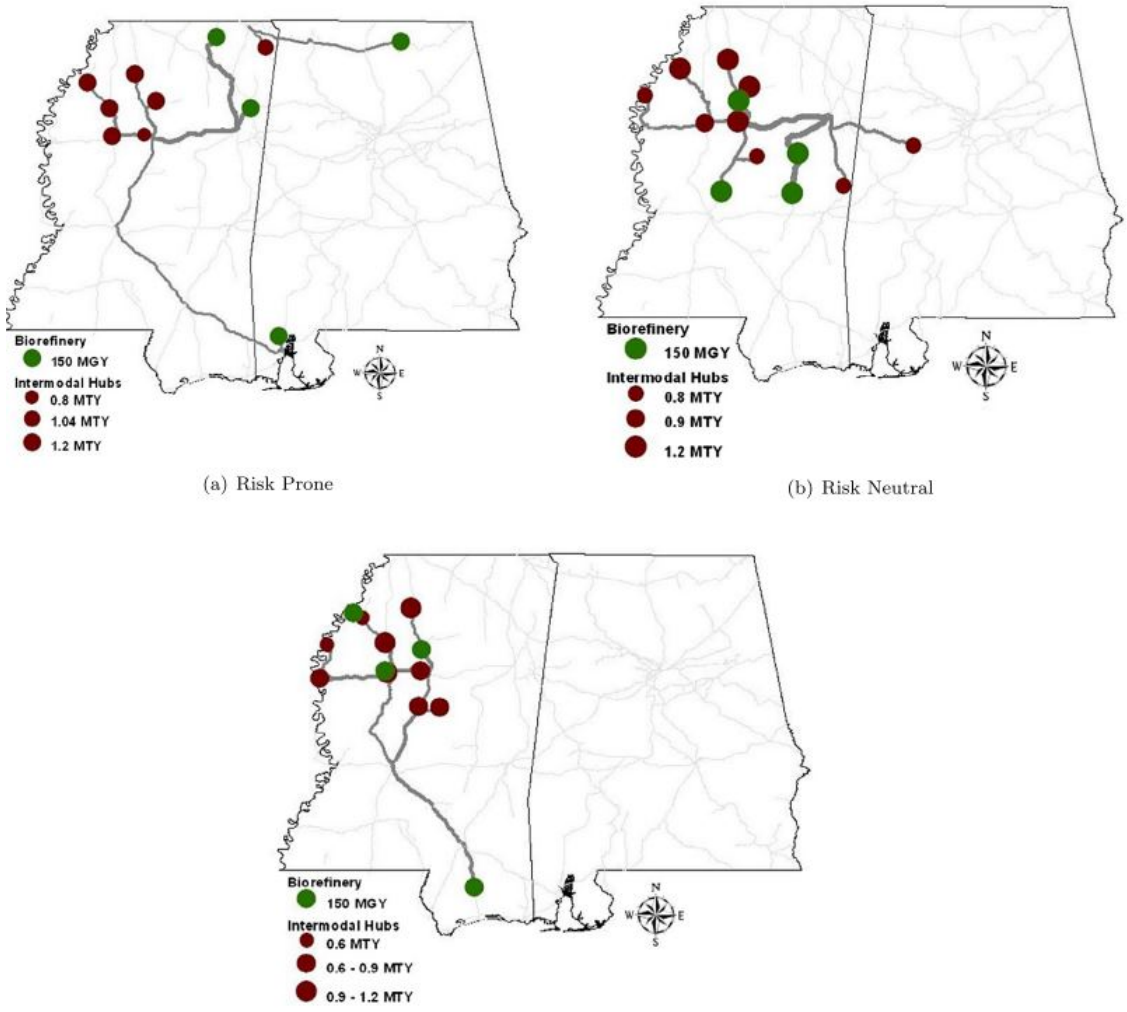
the number of containers transported between the multi-modal facilities, and bio-fuel production under various budget scenarios. Note that the reliability improvement cost for a given link is a function of several important parameters, such as the number of bridges within the link, the existing condition of the repair and renovation of a railroad track, and the proximity of that railroad track to lake, cannel, and river. These parameters as well as the total reliability improvement cost, for a given link varies. Sensitivity analyses are conducted to determine if the estimated reliability improvement cost changes and how that change impacts the bio-fuel supply chain network. Figure 2.9 shows that with the increase in unit reliability improvement cost, the bio-fuel supply chain network is more stable under $10 million budget case compared to $5 million budget case. If the estimated unit reliability improvement cost increases more than 10%, then the unit bio-fuel cost increases to $3.82/gallon from $3.58/gallon in the $5 million budget scenario. However, this cost only increases to $3.64/gallon from $3.58/gallon for the same experimental conditions in the $10 million budget case because the decision maker can improve the reliability of several links, even though the reliability improvement cost is high. The decision to improve impacts the selection of number of facilities operating in a given time period and eventually impacts the number of containers transported between the multi-modal facilities, as shown in the second and third sub-figures in Figure 2.9. Therefore, the biomass, which was initially supposed to be shipped through multi-modal facilities, is diverted to highways by trucks because of the high risk associated with multi-modal facility-to-facility links. This not only increases the unit transportation cost for biomass, but it also increases the unit bio-

fuel delivery cost. In summary, the size of the budget impacts the reliability improvement cost, which in turn affects the bio-fuel supply chain network.

Impact of different risk levels on system performance: We now analyze the impacts of different risk levels on bio-fuel supply chain network. Figure 2.10 shows the network representation under three different risk levels: (a) risk prone, (b) risk neutral, and (c) risk averse. We made the following definitions for this three different risk levels:

Definition 1 *Assume that link* (j,k) *can be divided into* n_{jk} *sections, each of which has a failure probability of* q_ℓ *for* $\ell = 1, \ldots, n_{jk}$*. Thus, a risk averse decision maker will estimate the link failure probability* q_{jk} *as* $q_{jk} = \max_{\ell=1,\ldots,n_{jk}} q_\ell$*. Similarly, a risk prone decision maker will estimate the link failure probability* q_{jk} *as* $q_{jk} = \min_{\ell=1,\ldots,n_{jk}} q_\ell$ *and a risk neutral decision maker will estimate the link failure probability* q_{jk} *as* $q_{jk} = avg_{\ell=1,\ldots,n_{jk}} q_\ell$

Clearly, when the decision maker adopts a risk prone attitude, then the model selects bio-refineries further away from the multi-modal facilities (Figure 2.10(a)). However, when the decision maker adopts a risk averse attitude, the model selects most of the bio-refineries in Northwest Mississippi that are close to their multi-modal facilities (Figure 2.10(c)). The reason behind selecting a number of multi-modal facilities and bio-refineries in the Northwest is this region produces abundance of biomass and usually is considered as a less disaster prone area (see Figure 2.3a and 2.4). Note that these experiments were conducted under $5 million budget case. In summary, depending on adopting different risk levels it impacts the bio-fuel supply chain network.



(a) Risk Prone

(b) Risk Neutral

(c) Risk Averse

Figure 2.10

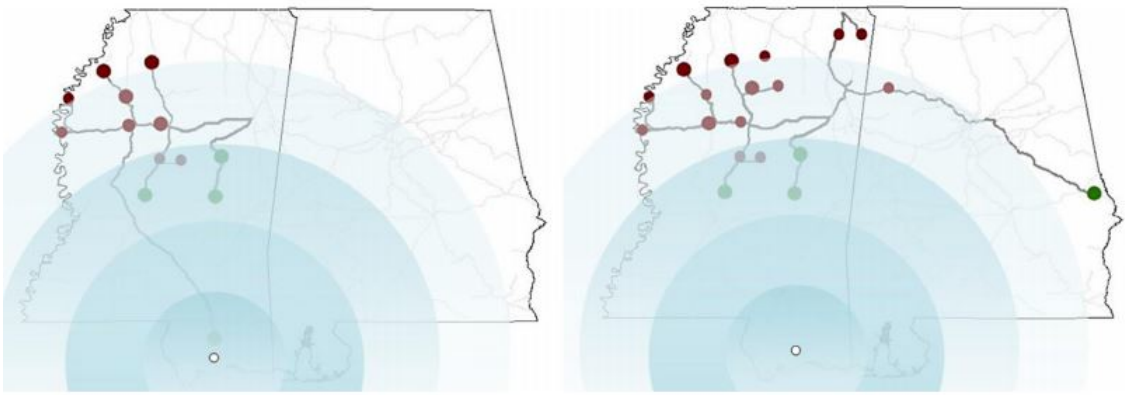
Impact of different risk levels on system performance

Impact of biomass supply changes on system performance: Table 2.2 presents the impact of supply changes on the unit cost of bio-fuel, the number of multi-modal facilities, the number of containers transported between the multi-modal facilities, and bio-fuel production. The results indicate that a 5% increase in biomass supply reduces the unit cost of bio-fuel by 11.08%. Therefore, the unit cost of bio-fuel lessened to $3.16 per gallon from $3.56 per gallon as biomass supply increased by 5%. A 10% increase in biomass supply decreases the unit cost of bio-fuel by 19.43% ($2.87 per gallon). Furthermore, the number of multi-modal facilities and the number of containers transported between the facilities are impacted by the biomass supply changes. For example, a 10% increase in biomass supply increases the number of multi-modal facilities by 40% to handle the increase of biomass shipped. This results in an additional 8.54% increase in containers transported between the multi-modal facilities. With the changes in the number of multi-modal facilities and the containers transported between them, the biomass supply will also impact the bio-fuel production, as shown in Table 2.2.

Table 2.2

Impact of supply changes on system performance

	Supply changes				
Item	-10%	-5%	0%	+5%	+10%
Unit cost of bio-fuel ($/gallon)	4.40	3.95	3.56	3.16	2.87
Number of multi-modal facilities	8	10	10	11	14
Number of containers transported	61,659	65,086	68,512	71,936	74,361
Bio-fuel produced (MGY)	447.64	472.51	497.38	522.25	547.12



(a) Budget = $5 million (b) Budget = $10 million

Figure 2.11

Identifying different hurricane scenarios in the bio-fuel supply chain network

Impact of different disruption scenarios on system performance: To quantify the benefits of designing a reliable bio-fuel supply chain system, we create different realistic hurricane scenarios under two different budget considerations: (a) $5 million and (b) $10 million. The reliability improvement suggested by model **[RM]** is tested by simulating hurricanes on the coastal regions of Mississippi and Alabama, generally considered as highly probable locations for such natural disasters to occur. Assume that the multi-modal facility connecting links located near the epicenter of the hurricane are more prone to disaster than those located toward the outer surface. Four different circles from the epicenter of the hurricane are developed to represent the failure probabilities of the transportation links and facilities in those regions. The darker the color of the circles indicate a higher chance of failure for the facilities within the region. Assume that there is a 50% chance that the transportation links located within the first circle will fail because the epicenter of the hurricane is located in the first circle. Though the radius of the circles will vary based

on the strength of the hurricane's strikes, this study sets the radius at 60 miles by assuming that a medium hurricane strikes the coast. Every 60 mile interval from the epicenter of the hurricane, the failure probability of the transportation links drops down to 15%, 10%, and 5%, respectively. These scenarios are illustrated in Figure 2.11.

We now solve model **[RM]** using these disruption scenarios and compare the solutions with the minimum cost model solutions. Note that a minimum cost model does not consider disruption and can be obtained by setting $\{q_{jt}\}_{j\in\mathcal{J},k\in\mathcal{K}} = \{\Delta_{jk}\}_{j\in\mathcal{J},k\in\mathcal{K}} = 0.0$ into the objective function value of model **[RM]**. Under normal conditions, the minimum cost model provides the minimum delivery cost, \$3.44/gallon, compared to the reliable cost model solution, \$3.53/gallon. However, under disaster scenarios, the reliable model solution outperforms the minimum cost model solution. We vary the failure probability of the circles by $\pm 5\%$ from the base scenario, and observe that the minimum cost model solution varies between \$3.85-\$3.96/gallon, compared to \$3.68-\$3.76/gallon for the \$5 million budget scenario and \$3.61-\$3.69/gallon for the \$10 million budget case. Thus, the decision of allocating more budget is justified to improve the reliability of the transportation network.

Algorithmic performance: This section presents the computational experience of solving model **[RM]** using the algorithm proposed in Section 3.2. We implemented the generalized Benders algorithm on the same version of GAMS, where we used CPLEX for solving the mixed-integer linear programming master problem **[BD(M)]** and CONOPT for solving the nonlinear subproblem **[BD(SUB)]**. The generalized Benders decomposition algorithm is terminated when at least one of the following conditions is met: (a) the

optimality gap ($\epsilon = |UB - LB|/UB$) falls below a threshold value $\epsilon = 0.01$; (b) the maximum time limit $t_{max} = 10,800$ CPU seconds is reached; or (c) the maximum iteration limit $Iter_{max} = 1,000$ is reached. Table 4.3 summarizes the problem size considered for analyzing the performance of the solution algorithms. The table further reports the problem size considered before and after the processed network. Ideally, model **[RM]** should account a fully connected network where all the source nodes are connected with the destination nodes (shown in case 1 and 2 in Table 4.3). However, having too many integer variables significantly increases the computational complexity of the problem. We then reduced the potential links to only those that are likely to be in an optimal solution. We performed the following preprocessing techniques to reduce the complexity of the original network. The number of variables and constraints reported in Table 4.3 (Case 3 and 4) is a result of this preprocessing.

- A study conducted by Brower [11] reported that moving biomass through trucks more than 50 miles to a feedstock supply point severely impacts profits. We use this information to remove all the arcs greater than 50 miles that connects a feedstock supplier $i \in \mathcal{I}$ with a multi-modal facility $j \in \mathcal{J}$.
- Multi-modal facilities are typically used for transporting high volume traffics at longer distances. Therefore, we remove all the arcs between two multi-modal facilities if the distance between them is smaller than 20 miles.
- We assume that the bio-fuel will not transport through trucks from a bio-refinery to a demand city if they are located 200 miles away. Therefore, we remove all the arcs greater than 200 miles that connects a bio-refinery $k \in \mathcal{K}$ with a demand city $g \in \mathcal{G}$.

Table 2.3

Experimental problem sizes

Cases	$\|\mathcal{J}\|$	$\|\mathcal{I}\|$	$\|\mathcal{K}\|$	$\|\mathcal{G}\|$	$\|\mathcal{L}\|$	Binary Variables	Integer Variables	Continuous Variables	Total Variables	Total Constraints
Before Preprocessing										
1	45	99	36	60	5	405	1,620	328,344	330,369	488,194
2	71	99	36	60	5	535	2,556	514,608	517,699	769,982
After Preprocessing										
3	45	99	36	60	5	405	1,140	146,259	147,804	216,175
4	71	99	36	60	5	535	1,807	279,311	281,653	381,486

Table 2.4

Data parameter settings for algorithm comparison

Data Set #	Cases	q_{jk}	Budget (mil. $)
S1	3	U[0.05, 0.25]	5.0
S2	3	U[0.05, 0.25]	10.0
S3	4	U[0.05, 0.25]	5.0
S4	4	U[0.05, 0.25]	10.0

To compare the performance between GAMS/BONMIN and generalized Benders decomposition algorithm, we use six different data sets where four data sets are presented in Table 2.4 and the two remaining sets include data obtained from the real world. Note that the failure probability between the links of the multi-modal facilities for data sets S1-S4 are generated using an uniform distribution, i.e., $q_{jk} \sim U[0.05, 0.25]$. Finally, Table 2.5 compares the computational results in solving model **[RM]** between GAMS/BONMIN and the generalized Benders decomposition algorithm. Note that the algorithm is enhanced with the integer cuts, knapsack inequalities, and logistics constraints discussed in section 2.4.2. The best result is identified for each problem solved and presented in boldface. Since a stopping criteria for the algorithm is $\epsilon \leq 1.0\%$, the algorithms are stopped when such a

solution is found within the maximum time limit. The algorithm that gave the smallest running time is highlighted. Otherwise, if such a quality solution is not found within the maximum time limit or number of iterations, the algorithm with the smallest optimality gap is highlighted. Based on the results presented in Table 2.5, it is observed that the generalized Benders decomposition algorithm is capable of solving all the problem instances (22 out of 22) in less than 1% optimality gap on a reasonable amount of time as compared to GAMS/BONMIN. For the instances that are interrupted because of the time limit, we calculate the average gap at the end of the time limit and label them in column "Avg Gap". In overall, the average optimality gap and running time for generalized Benders decomposition algorithm is reported as 0.89% and 803.1 CPU seconds compared to 48.41% and 10,800 CPU seconds as reported by GAMS/BONMIN.

Table 2.5

Computational performance of the solution algorithm

Data Set	GAMS/BONMIN			Generalized Benders			
	# of Solved	Avg Gap (%)	Avg CPU (sec)	# of Solved	Avg Gap (%)	Avg CPU (sec)	Avg Iter
Real-1[a]	0/1	47.86	10,800	1/1	**0.87**	**872.9**	72
Real-2[b]	0/1	49.11	10,800	1/1	**0.95**	**859.4**	71
S1	0/5	42.18	10,800	5/5	**0.86**	**611.7**	58
S2	0/5	41.04	10,800	5/5	**0.83**	**645.7**	61
S3	0/5	52.64	10,800	5/5	**0.91**	**923.4**	77
S4	0/5	57.67	10,800	5/5	**0.94**	**905.5**	75
Average	0/22	48.41	10,800	22/22	**0.89**	**803.1**	69

[a]Real failure probability data with \$5.0 million budget and $|\mathcal{J}| = 71$
[b]Real failure probability data with \$10.0 million budget and $|\mathcal{J}| = 71$

2.6 Conclusion

This paper studies a pre-disaster planning problem that seeks to strengthen the links between the multi-modal facilities of a bio-fuel supply chain network. A mixed-integer nonlinear programming model **[RM]** is developed to determine the optimal locations for multi-modal facilities and bio-refineries, offer suggestions on reliability improvement at vulnerable links, production, and make transportation decisions under both normal and disrupted scenarios. The aim is to assist investors in determining which links' reliability can be improved under specific budget limitations so that the bio-fuel supply chain network can prevent possible losses when transportation links are disrupted because of natural disasters. The failure probability of the links between the multi-modal facilities is estimated with a spatial statistics model developed from real world data. Finally, this paper presents a generalized Benders decomposition algorithm for solving our reliable, pre-disaster planning problem **[RM]**. This enhanced approach incorporates several algorithmic improvements such as integer cuts, knapsack inequalities, and logistics constraints. Computational results show that the enhanced generalized Benders decomposition algorithm is capable of solving realistic size problem instances efficiently compared to GAMS/BONMIN.

The states of Mississippi and Alabama are used as a testing ground for this model, and several experiments provide insights about the advantages of using model **[RM]** in a bio-fuel supply chain network. As part of the numerical experimentation, some realistic hurricane scenarios are presented to determine the potential impact that pre-investing may have on improving the bio-fuel supply chain network's reliability on vulnerable transportation links considering limited budget availability. In case of a disaster, the unit bio-fuel

delivery cost provided by the minimum cost model increases to $3.96/gallon, compared to $3.69/gallon provided by the reliable model solution. Moreover, the impacts of several features of the bio-fuel supply chain were examined, including network budget allocation, reliability improvement costs, the risk level for decision makers, and biomass supply changes. The findings of these experiments, presented here, will help decision makers design reliable logistics networks for the bio-fuel supply chain network.

This work can be extended in several directions. This study ignores the impacts of seasonality of biomass, an important factor to consider in bio-fuel supply chain literature. Our work can further be extended by incorporating into the modeling process the impacts of traffic congestion caused by natural disasters. Furthermore, more realistic and correlated multi-modal facility failure patterns should be incorporated into the modeling framework. Finally, research is required to generate stronger cuts that will further enhance the Generalized Benders decomposition algorithm. These issues will be addressed in future studies.

CHAPTER 3

A HYBRID DECOMPOSITION ALGORITHM FOR DESIGNING A MULTI-MODAL TRANSPORTATION NETWORK UNDER BIOMASS SUPPLY UNCERTAINTY

Co-firing biomass with coal has been receiving extensive attention from the energy community due to the several added benefits of the process. First, co-firing biomass reduces green house gas (GHG) emissions due to its displacing coal with biomass [26]. Second, biomass co-firing can be accomplished using the existing coal-fired power plant infrastructures. Therefore, no additional capital investment is required. Third, co-firing minimizes waste such as wood and agricultural waste and the environmental problem associated with its disposal [100]. On the contrary, few studies [57, 100], address some limitations of co-firing biomass with coal: (1) co-firing biomass reduces boiler efficiency and eventually the overall systems efficiency; (2) biomass has lower heating and density values, both of which are undesirable from a production point of view; and, (3) biomass supply is dispersed geographically, adding logistical challenges. This study explores some of the logistical challenges in the biomass supply chain system that will help the managers take better decisions.

The efficiency of co-firing biomass depends on how economically the biomass is transported from the feedstock supply points to the coal plants. This process necessitates devel-

oping an integrated logistics network that efficiently connects origin fields to destination coal plants with proper transportation modes. However, designing such a network is challenging due to the physical characteristics of biomass. Biomass is bulky and difficult to transport; its supply is highly seasonal and uncertain; and biomass is widely dispersed geographically. Therefore, modes such as rail can be considered as the viable modes of transportation to deliver biomass from feedstock suppliers to coal plants. However, the operations in multi-modal facilities can be highly impacted by the seasonality of biomass. For instance, the harvesting season for corn stover follows the harvesting of corn which starts from early September and ends in November. On the other hand, woody biomass is available all year except three months in the winter. In addition to the feedstock seasonality, it is observed that the biomass supply fluctuates extensively from one year to another. This statement is supported by Figure 3.13 which shows the fluctuation of cornstover availability in the four states Mississippi (MS), Alabama (AL), Tennessee (TN), and Louisiana (LA) in Southern United States over 1980-2014 [111]. The figure shows that the availability of biomass cannot be predicted accurately. The biomass supply fluctuates highly from one time period to another depending on climatic conditions such as rain, temperature, and humidity along with other extreme events such as, natural disasters and human intervention [90]. This poses a significant logistics challenge in the network and highlights the need for developing optimization models to design a cost-efficient, reliable biomass cofiring supply chain under feedstock supply uncertainty. The study of the biomass supply chain optimization has increased rapidly in recent years. We group the existing literatures into deterministic and stochastic research to reflect the contribution of our modeling and

algorithmic approaches used in this paper. Studies conducted by Zamboni et al. [132], Eksioglu et al. [29], Eksioglu et al. [30], Huang et al. [49], An et al. [1], Bai et al. [5], Walther et al. [120], Xie and Ouyang [129], Zhang et al. [3], Roni et al. [100], Xie et al. [127] and Memisoglu and Uster [80] analyze plant location and transportation issues in biofuel supply chain networks under a deterministic setting. These studies are further extended to consider a case when the disruption in bio-refineries (e.g., Li et al. [66], Wang and Ouyang [122], Bai et al. [6]) or in multi-modal facilities (e.g., Marufuzzaman et al. [76], Marufuzzaman and Eksioglu [73]) or in the links between the multi-modal facilities (e.g., Poudel et al. [92]) impact the biofuel supply chain network. Another stream of research in bio-fuel supply chain community develops multi-objective optimization model that not only minimizes cost but also minimizes Green House Gas (GHG) emissions from the supply chain network (e.g., Zamboni et al. [131], Giarola et al. [41], You et al. [130]).

All the studies discussed above assume that the model input parameters (e.g., biomass supply, demand, technology) are known and thus fail to capture the impact of system uncertainties in the biofuel supply chain network configuration. To represent a more realistic case, studies conducted by Cundiff et al. [23], Kim et al. [59], Chen and Fan [17], Gebreslassie et al. [38], Awudu and Zhang [3], and Marufuzzaman et al. [75] account uncertainties in biomass supply, demand, technology, and pricing. A brief overview of considering uncertainty and sustainability in a biofuel supply chain network can be found from a recent study by Awudu and Zhang [54]. Although both the deterministic and stochastic models have done a great job in capturing the overall system level design, the models did not explicitly consider the option of using long haul transportation modes such as rail

or barge as a viable mode of transportation to carry biomass from multi-modal facilities to bio-refineries except the studies from Marufuzzaman et al. [76], Roni et al. [100], Marufuzzaman and Eksioglu [73]. Introducing multi-modal facilities in the bio-fuel supply chain network not only reduces cost but also alleviates congestion in the highways and as a result improves safety. The case became much stronger after Idaho National Laboratory proposed a biomass delivery system that preprocess biomass prior to transporting [48]. Densified biomass has physical characteristics which are similar to corn, soybean and other grains; therefore, it is easy to load/unload and transport, and thus long hauls become an option. Recent studies by Marufuzzaman et al. [76], Roni et al. [100], Marufuzzaman and Eksioglu [73], [94] captured this specific product information in the modeling framework and identified the viability of using multi-modal facilities in a bio-fuel supply chain network. Our model can be considered as a direct extension of these studies in the sense that we captured feedstock uncertainty in a multi-time period problem to transport biomass from feedstock supply sites to coal-plants.

Although stochastic programming approaches provide more reliable solutions, they are often achieved with high computational burden. Commercial solvers fail to solve the model directly for even a small set of scenarios for most of the real life problems. Therefore, decomposition algorithms are often employed to overcome the computational difficulties. Note that our problem can be considered as an extension of multi-time period facility location problem that handles multi-commodities under uncertainty. Until now a number of decomposition techniques have been developed to solve such stochastic facility location problems. Schutz et al. [103] develops a hybrid decomposition algorithm that uses a sam-

ple average approximation algorithm along with a dual decomposition algorithm to solve a Norwegian meat supply chain problem under uncertainty. Santoso et al. [102] develops a hybrid decomposition algorithm that uses sample average approximation algorithm and an enhanced Benders decomposition algorithm to solve a realistic size supply chain network designing problem under uncertainty. Rajgopal et al. [95] proposes a double-cut L-shaped algorithm to solve a two-stage stochastic supply chain network designing problem under uncertainty. Keyvanshokooh et al. [58] develops an accelerated Benders decomposition algorithm to solve a closed-loop supply chain network design problem under uncertainty. Despite all the efforts, little work has been done so far to tackle a fairly realistic size stochastic bio-fuel supply chain network designing problem. Work done by Chen and Fan [17] and Huang et al. [50] uses Progressive Hedging Algorithm (PHA) to solve a single and multi-time period two-stage stochastic programming model formulation under feedstock supply uncertainty. The authors are able to solve up to eight feedstock supply scenarios within 1% optimality gap. Marufuzzaman et al. [75] develops a Lagrangian relaxation based multi-cut L-shaped algorithm to solve a two-stage stochastic bio-crude supply chain network problem of up to ten feedstock supply scenarios. Gebreslassie et al. [38] solves a single time period bio-refinery supply chain network problem using a multi-cut L-shaped algorithm of up to one thousand feedstock supply scenarios. In this study, we extend the literatures by generating realistic scenarios obtained from prediction errors of the historical and forecasted biomass supply availabilities and then develop a hybrid solution approach that combines Sample average approximation algorithm with an enhanced Progressive hedging algorithm to solve a multi-commodity, multi-time period biomass co-

firing supply chain network designing problem of up to 500 scenarios. To enhance the Progressive hedging algorithm, we use several stronger techniques, including local and global adjustment strategies [19], dynamic penalty parameter updating techniques, and Rolling horizon heuristics. The proposed algorithm is then validated via a real-world case study with data from Mississippi and Alabama.

In summary, the major contribution of this research to the existing biomass supply chain literatures are as follows:

- This study proposes a two-stage stochastic mixed-integer linear programming model formulation for the design and management of a biomass co-firing supply chain problem under feedstock supply uncertainty. The output of our model provides season-wise utilization of multi-modal facilities, number of containers transported between the multi-modal facilities, and the amount of biomass processed, stored, and transported from multiple feedstock supply sites to coal plants under biomass supply uncertainty. Our model extends prior biomass co-firing optimization models (e.g., Roni et al. [100], Karimi et al. [57], [31]) by incorporating seasonality and uncertainty in the modeling framework.
- We generate realistic scenarios obtained from prediction errors of the historical and forecasted biomass supply availabilities. The scenarios are then used in the stochastic program to generate more useful and realistic solutions. Such approach has not been addressed yet by any literatures in the biomass supply chain community.
- We develop a hybrid solution approach that combines Sample average approximation with an enhanced Progressive hedging algorithm to solve a multi-commodity,

multi-time period biomass co-firing supply chain network designing problem. To enhance the Progressive hedging algorithm, we use several stronger techniques, including local and global adjustment strategies [14], dynamic penalty parameter updating techniques, and Rolling horizon heuristics. Computational results confirm the efficiency and robustness of the proposed solution method.

- Finally, we validate the modeling results by developing a case study using data from Mississippi and Alabama. The outcome of this study provides a number of managerial insights such as deployment of multi-modal facilities, storage, processing and transportation of biomass under different biomass supply variability levels. These insights can effectively aid decision makers to design a reliable biomass co-firing supply chain network. Finally, we demonstrated how the mean biomass supply changes impacts the supply chain network configuration and also justified the use of considering stochastic programming modeling approach over deterministic approach.

The remainder of this paper is organized as follows. Section 4.2 presents the model formulation. Section 3.2 discusses how to generate scenario trees and discusses a hybrid decomposition algorithm that combines Sample average approximation algorithm with Progressive hedging algorithm to efficiently solve the proposed model. Section 4.4 conducts numerical experiments to verify the algorithm performance and to draw managerial insights. Section 4.5 concludes this paper and discusses future research directions.

3.1 Problem Description and Model Formulation

This section presents a two-stage stochastic programming model formulation for the design and management of a biomass co-firing supply chain network while taking stochastic nature of biomass supply into account. In the two-stage stochastic programming model formulation, the first-stage decides the location, capacity, and timing to use a multi-modal facility among candidate locations prior to realization of any random events. After the first-stage decisions are made, the random events are realized and the second-stage decisions such as the cost of procuring, storing, and transporting biomass from feedstock supply sites to coal plants are made. The objective is to minimize the first-stage and expected second-stage costs across all possible feedstock supply scenarios for the biomass co-firing supply chain network. Figure 4.3 gives an example of a logistics network for co-firing biomass with coal that consists of two feedstock suppliers, three multi-modal facilities, and one coal plant. One can view this network as a set of nodes and arcs where nodes represent locations where one could potentially use a multi-modal facility, or represent harvesting sites that supply biomass feedstock, or coal plants where the biomass are to be delivered. Arcs that connect harvesting sites with multi-modal facilities and coal plants represent transportation arcs. Finally, the arcs that connect same facility in two consecutive time periods are the inventory arcs. The problem is to minimize the overall planning and operational cost where the planning decisions (e.g., use of multi-modal facilities) are made at the beginning of the year while the operational decisions (e.g., transportation, storage, processing) are time dependent and made for every season over a year.

Let $\mathbb{G}(\mathcal{N}, \mathcal{A})$ denote a logistics network where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of arcs. Set $\mathcal{N}$ consists of the set of harvesting sites $\mathcal{I}$, the set of candidate multi-modal facilities $\mathcal{J}$, and the set of coal plants $\mathcal{K}$, i.e., $\mathcal{N} = \mathcal{I} \cup \mathcal{J} \cup \mathcal{K}$. The set of arcs $\mathcal{A}$ is partitioned into three disjoint subsets, i.e., $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$, where $\mathcal{A}_1$ represents the set of arcs joining harvesting sites $\mathcal{I}$ with multi-modal facilities $\mathcal{J}$; $\mathcal{A}_2$ represents the set of arcs between multi-modal facilities $\mathcal{J}$ and coal-plants $\mathcal{K}$; and finally, $\mathcal{A}_3$ represents the set of arcs that directly connect from harvesting sites $\mathcal{I}$ to coal plants $\mathcal{K}$. Let $\mathcal{B}$ denote the set of feedstock types to be distributed in the logistics network in different time periods of a year, and let $\mathcal{T}$ denotes the set of time periods to be considered in this problem. We further denote Ω as the sample space of the random event, where $\omega \in \Omega$ defines a particular realization. Let $s_{bit\omega}$ denote the amount of biomass of type $b \in \mathcal{B}$ available at site $i \in \mathcal{I}$ in time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$. The model considers fixed number of scenarios $(|\Omega|)$ each of which represents a particular realization of biomass supply and there is a positive probability ρ_ω ($\rho_\omega \geq 0$) associated with each scenario $\omega \in \Omega$. Each coal plant $k \in \mathcal{K}$ demands d_{kt} tons of biomass in period $t \in \mathcal{T}$. We consider that the unmet demand of the biomass can be substituted. The per unit cost of not satisfying the demand at each coal plant $k \in \mathcal{K}$ in period $t \in \mathcal{T}$ is denoted by π_{kt}. This will be the threshold cost that customer want to pay for the biomass. If the unit delivery cost of biomass exceeds this threshold then it will be beneficial to use the substitution rather than getting biomass on higher cost.

For nodes $j \in \mathcal{J}$, ψ_{ljt} denotes the fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ in time period $t \in \mathcal{T}$. We choose the candidate location for multi-modal facilities

which already exist at given location $j \in \mathcal{J}$ and the fixed cost represents the additional cost to build the infrastructure (e.g., building additional track, purchasing lifts, track switch). We assume that the coal plants are co-located with multimodal facilities. We further assume that the transportation cost between multi-modal facilities to nearby coal plants is negligible. This assumption is made based on the fact that most of the coal plants have railway access since railway is the common mode to transport coal. Each arc $(i, j) \in \mathcal{A}_1$ carries biomass of type $b \in \mathcal{B}$ from a harvesting site i to a multi-modal facility j and are generally located close to each other (e.g., 10-20 miles). Therefore, trucks are preferred on $(i, j) \in \mathcal{A}_1$ and its unit transportation cost is represented by c_{bij}. The arcs $(j, k) \in \mathcal{A}_2$ carry large-volume long-haul traffic and use rail as a major transportation mode to transport biomass from multi-modal facilities to coal plants. We represent c_{bjkt} as a unit transportation cost along arc $(j, k) \in \mathcal{A}_2$ in time period $t \in \mathcal{T}$. Therefore, a unit flow along an origin-destination route $\{(i, j), (j, k)\}$ for biomass of type $b \in \mathcal{B}$ in time period $t \in \mathcal{T}$ costs $c_{bijkt} = c_{bijt} + c_{bjkt}$. Biomass is usually transported in cargo containers between the multi-modal facilities. Therefore, we consider a fixed cost ξ_{jkt} which represents the cost associated with loading and unloading biomass in the cargo containers in time period $t \in \mathcal{T}$. We further denote c_{bikt} as the unit cost of transporting biomass of type $b \in \mathcal{B}$ along arcs $(i, k) \in \mathcal{A}_3$ in time period $t \in \mathcal{T}$. These arcs are preferred when the feedstock sites are located close to the coal plants and direct shipments of biomass using trucks are considered cheaper compared to multi-modal facilities. Finally, biomass is required to be purchased from the farmers and collected and stored before hauling it on truck. Therefore, we denote δ_{bit} and γ_{bit} as the unit purchasing and storage cost for biomass of type $b \in \mathcal{B}$ at

location $i \in \mathcal{I}$ in time period $t \in \mathcal{T}$. Let us now introduce the following notation for our two-stage stochastic programming formulation:

Sets:

- $\mathcal{I}$: set of harvesting sites (farms)
- $\mathcal{J}$: set of multi-modal facilities
- $\mathcal{K}$: set of coal plants
- $\mathcal{L}$: set of capacities
- $\mathcal{T}$: set of time periods
- $\mathcal{B}$: set of biomass types (b_1 for corn-stover and b_2 for forest residues)
- Ω: set of scenarios

Parameters:

- ψ_{ljt}: fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$
- ξ_{jkt}: fixed cost of a cargo container for transporting biomass along arc $(j,k) \in \mathcal{A}_2$ in time period $t \in \mathcal{T}$
- c_{bikt}: unit transportation cost for biomass of type $b \in \mathcal{B}$ along arc $(i,k) \in \mathcal{A}_3$ in time period $t \in \mathcal{T}$
- c_{bijkt}: unit transportation cost for biomass of type $b \in \mathcal{B}$ along arc $(i,j,k) \in \mathcal{A}_1 \cup \mathcal{A}_2$ in time period $t \in \mathcal{T}$
- $s_{bit\omega}$: amount of biomass of type $b \in \mathcal{B}$ available at site $i \in \mathcal{I}$ in period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- d_{kt}: demand for biomass at coal plant $k \in \mathcal{K}$ in period $t \in \mathcal{T}$
- c^{cap}: capacity of cargo container
- h_k^{cap}: biomass holding capacity at coal plant $k \in \mathcal{K}$
- C_{lj}: biomass storage/handling capacity of a multi-modal facility of size $l \in \mathcal{L}$ at location $j \in \mathcal{J}$
- α_b: deterioration rate of biomass of type $b \in \mathcal{B}$

- p_{bkt}: unit biomass processing cost of type $b \in \mathcal{B}$ at coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$
- h_{bkt}: unit inventory holding cost for biomass of type $b \in \mathcal{B}$ at coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$
- δ_{bit}: unit biomass purchasing cost of type $b \in \mathcal{B}$ at location $i \in \mathcal{I}$ in time period $t \in \mathcal{T}$
- γ_{bit}: unit biomass storage cost of type $b \in \mathcal{B}$ at location $i \in \mathcal{I}$ in time period $t \in \mathcal{T}$
- π_{kt}: unit penalty cost of not satisfying demand at coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$
- ρ_{ω}: probability of scenario $\omega \in \Omega$

Decision Variables:

- Y_{ljt}: 1 if a multi-modal facility $j \in \mathcal{J}$ of capacity $l \in \mathcal{L}$ is used at time period $t \in \mathcal{T}$; 0 otherwise
- $Z_{jkt\omega}$: number of cargo containers transported between multi-modal facility $j \in \mathcal{J}$ to coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $X_{bikt\omega}$: amount of biomass of type $b \in \mathcal{B}$ transported from harvesting site $i \in \mathcal{I}$ to coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $X_{bijkt\omega}$: amount of biomass of type $b \in \mathcal{B}$ transported from harvesting site $i \in \mathcal{I}$ to coal plant $k \in \mathcal{K}$ through multi-modal facility $j \in \mathcal{J}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $P_{bkt\omega}$: amount of biomass of type $b \in \mathcal{B}$ processed in coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $H_{bkt\omega}$: amount of inventory of type $b \in \mathcal{B}$ stored in coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $U_{kt\omega}$: amount of biomass shortage at coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$

We now introduce the following first and second-stage decision variables for our two-stage stochastic programming model formulation. The first-stage decision variables $\mathbf{Y} := \{Y_{ljt}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ determine the size, location, and time to use a multi-modal facilities, i.e.,

$$Y_{ljt} = \begin{cases} 1 & \text{if a multi-modal facility of size } l \text{ is used at location } j \text{ at time period } t \\ 0 & \text{otherwise;} \end{cases}$$

The second-stage decision variables $\mathbf{Z} := \{Z_{jkt\omega}\}_{j\in\mathcal{J},k\in\mathcal{K},t\in\mathcal{T},\omega\in\Omega}$ decide the number of cargo containers transported between multi-modal facility $j \in \mathcal{J}$ to coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$; $\mathbf{X} := \{X_{bert\omega}\}_{b\in\mathcal{B},(e,r)\in\mathcal{A},t\in\mathcal{T},\omega\in\Omega}$ denote the flow of biomass of type $b \in \mathcal{B}$ along each link $(e, r) \in \mathcal{A}$ of the network at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$; $\mathbf{P} := \{P_{bkt\omega}\}_{b\in\mathcal{B},k\in\mathcal{K},t\in\mathcal{T},\omega\in\Omega}$ denote the amount of biomass of type $b \in \mathcal{B}$ processed in coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$; $\mathbf{H} := \{H_{bkt\omega}\}_{b\in\mathcal{B},k\in\mathcal{K},t\in\mathcal{T},\omega\in\Omega}$ denote the amount of inventory of type $b \in \mathcal{B}$ stored in coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$; and $\mathbf{U} := \{U_{kt\omega}\}_{k\in\mathcal{K},t\in\mathcal{T},\omega\in\Omega}$ denote the amount of biomass shortage at coal plant $k \in \mathcal{K}$ in period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$.

The objective function of the biomass co-firing supply chain network model **[CSC]** minimizes the first-stage costs and the expected value of the random second-stage costs. The first-stage decisions include the fixed costs of using multi-modal facilities at different time period of a year. These decisions are required to be made before the uncertainty is revealed. However, after the uncertainty is revealed the second stage decisions are made which include transportation, storage, processing, and penalty costs. The following is a

two-stage mixed-integer linear programming (MILP) model formulation of the problem referred to as model **[CSC]**.

$$\textbf{[CSC]}\ \underset{\mathbf{Y}}{Minimize} \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}\psi_{ljt}Y_{ljt} + \sum_{\omega\in\Omega}\rho_{\omega}\mathbb{Q}(\mathbf{Y},\omega) \tag{3.1}$$

Subject to

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

$$Y_{ljt} \in \mathbb{B} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T} \tag{3.3}$$

with $\mathbb{Q}(\mathbf{Y},\omega)$ being the solution of the following second-stage problem:

$$\begin{aligned}\mathbb{Q}(\mathbf{Y},\omega) = \underset{\mathbf{Z,X,P,H,U}}{Minimize} \sum_{t\in\mathcal{T}}\Bigg(&\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt\omega} + \sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}(c_{bikt}+\delta_{bit}\gamma_{bit})X_{bikt\omega}\\ &+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}(c_{bijkt}+\delta_{bit}+\gamma_{bit})X_{bijkt\omega} + \sum_{k\in\mathcal{K}}p_{bkt}P_{bkt\omega} + \sum_{k\in\mathcal{K}}h_{bkt}H_{bkt\omega}\Bigg)\\ &+\sum_{k\in\mathcal{K}}\pi_{kt}U_{kt\omega}\Bigg)\end{aligned} \tag{3.4}$$

Subject to

$$\sum_{k\in\mathcal{K}} X_{biktω} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} X_{bijktω} \leq s_{bitω} \quad \forall b\in\mathcal{B}, i\in\mathcal{I}, t\in\mathcal{T}, ω\in\Omega \tag{3.5}$$

$$\sum_{i\in\mathcal{I}} X_{biktω} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} X_{bijktω} + (1-\alpha_b) H_{b,k,t-1,ω} = H_{bktω} + P_{bktω} \quad \forall b\in\mathcal{B}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{3.6}$$

$$\sum_{b\in\mathcal{B}} P_{bktω} + U_{ktω} = d_{kt} \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{3.7}$$

$$\sum_{b\in\mathcal{B}} H_{bktω} \leq h_k^{cap} \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{3.8}$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}} X_{bijktω} \leq c^{cap} Z_{jktω} \quad \forall (j,k)\in\mathcal{A}_2, t\in\mathcal{T}, ω\in\Omega \tag{3.9}$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijktω} \leq \sum_{l\in\mathcal{L}} C_{lj} Y_{ljt} \quad \forall j\in\mathcal{J}, t\in\mathcal{T}, ω\in\Omega \tag{3.10}$$

$$Z_{jktω} \leq \sum_{l\in\mathcal{L}} \left\lceil \frac{C_{lj}}{c^{cap}} \right\rceil Y_{ljt} \quad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{3.11}$$

$$Z_{jktω} \in \mathbb{Z}^+ \quad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{3.12}$$

$$X_{bijktω}, X_{biktω}, H_{bktω}, P_{bktω}, U_{ktω} \in \mathbb{R}^+ \ \forall b\in\mathcal{B}, i\in\mathcal{I}, j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{3.13}$$

The objective function (3.1) is the sum of the first-stage cost and the expected second-stage cost over all scenarios. The first-stage cost represents the costs of using multi-modal facilities of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$ in period $t \in \mathcal{T}$. Constraints (4.2) indicate that at most one multi-modal facility of capacity $l \in \mathcal{L}$ is operating in a given location $j \in \mathcal{J}$ at time period $t \in \mathcal{T}$. Constraints (4.3) define binary restrictions for the first-stage decision variables. The second-stage objective function (4.4) consists of six parts: the first term represents the fixed cost of transporting cargo containers between the multi-modal

facilities, the second and third terms represent the expected transportation cost of biomass routed through highways and multi-modal facilities, the fourth and fifth terms represent the expected processing and holding cost of biomass at the coal plants, and the last term of the objective function represents the expected penalty cost for unsatisfied demand. Constraints (4.5) indicate that the amount of biomass of type $b \in \mathcal{B}$ shipped from a harvesting site $i \in \mathcal{I}$ at time period $t \in \mathcal{T}$ and under scenario $\omega \in \Omega$ is limited by its availability. Constraints (4.6) are the flow balance constraints which ensure that at any time period $t \in \mathcal{T}$ and under scenario $\omega \in \Omega$ the biomass of type $b \in \mathcal{B}$ can be either processed or stored in a coal plant $k \in \mathcal{K}$. Constraints (4.7) indicate that the demand of biomass will be fulfilled either through the processed biomass obtained from the distribution network or through substitute products available in the market. Constraints (4.8) limit the amount of biomass that can be stored in a coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ and under scenario $\omega \in \Omega$. Constraints (4.9) set a limit on the amount of biomass to be routed on arcs $(j, k) \in \mathcal{A}_2$ at time period $t \in \mathcal{T}$ and under scenario $\omega \in \Omega$. Constraints (4.10) indicate that the total amount of biomass shipped through multi-modal facility $j \in \mathcal{J}$ at time period $t \in \mathcal{T}$ and under scenario $\omega \in \Omega$ is limited by the facility capacity C_{lj}; $\forall l \in \mathcal{L}, j \in \mathcal{J}$. Constraints (4.11) serve as a valid-in-equality for model **[CSC]** which indicate that if no multi-modal facility is established at location $j \in \mathcal{J}$, then no containers will flow between arcs $(j, k) \in \mathcal{A}_2$ at time period $t \in \mathcal{T}$ and under scenario $\omega \in \Omega$. Finally, constraints (4.12) are the integrality constraints and (4.13) are the standard non-negativity constraints.

3.2 Solution Approach

By setting $|\Omega| = 1$ and $|\mathcal{T}| = 1$ i.e., a single scenario and a single season, we can show that problem **[CSC]** is a special case of a capacitated facility location problem which is known to be an $\mathcal{NP}$-hard problem [69]. Therefore, commercial solvers, such as CPLEX, cannot solve large-scale instances of this problem. In this section, we first use a prediction method to generate biomass supply scenarios and then use a hybrid decomposition algorithm that combines Sample average approximation algorithm and Progressive hedging algorithm to solve problem **[CSC]**. The goal is to generate high quality solution for problem **[CSC]** in a reasonable amount of time.

3.2.1 Scenario Generation for Biomass Supply Uncertainty

The availability of corn-stover varies significantly from one year to the other. Figure 3.13 supports this statement by showing Mississippi as a case example where the availability of corn-stover increases from 35 million tons in 2006 to 135 million tons in 2007 and then drops down to 98 million tons in 2008 [111]. Similar variations can be observed in other neighboring states of Mississippi, such as Alabama, Tennessee, and Louisiana (Figure 3.13). This mandates that there is a need to consider a large scenario set for biomass supply availability to develop a two-stage stochastic programming model formulation. However, such instances will not only increase the size of the problem but also will pose high computational challenges in solving model **[CSC]**. To alleviate this problem, a reliable prediction model needs to be developed that takes the historical biomass availability in a given region into account to predict the availability of biomass for the targeted year.

The prediction error provided by the model can be used to generate scenario trees for the two-stage stochastic programming model formulation. The aim is to reduce the scenario size for the two-stage stochastic programming model formulation.

Like Schutz et al., [103] and Keyvanshokooh et al., [58], we utilize a deterministic pth-order Vector Autoregressive Process (VAR) as the forecasting method and add a realization of error term $\epsilon^{\omega}_{\mathbf{bi,t+1}}$ to predicted the supply of corn-stover (b_1) in location $i \in \mathcal{I}$ at time $t+1$ under scenario $\omega \in \Omega$, denoted by $s^{\omega}_{bi,t+1}$:

$$\hat{\mathbf{s}}^{\omega}_{\mathbf{bi,t+1}} \quad = \quad \alpha + \sum_{m=1}^{p} \beta_m s_{bi,t+1-m} + \epsilon^{\omega}_{\mathbf{bi,t+1}} \tag{3.14}$$

where, $\hat{\mathbf{s}}^{\omega}_{\mathbf{bi,t+1}}$ denotes the $n \times 1$ vector of variables included in the VAR for year $t+1$; α is a constant parameter; β_i is an autoregressive parameter and there are p autoregressive matrices; and $s_{bi,t+1-m}$ is the historical biomass supply quantity in period $(t+1-m)$. We observed that the predictive performance of the first, third and fourth models are better than other (as shown in figure 3.1). The VAR(3) diagonal model seems to be the best and most parsimonious fit, so we used VAR(3) diagonal to fit our data and make a prediction. The Normalized Root Mean Square Error (NRMSE) for our pth-order VAR forecasting model is 0.2138.

The biomass supply scenario is now generated as follows:

$$\hat{\mathbf{s}}^{\omega}_{\mathbf{bi,t+j}} = \begin{cases} \alpha + \sum_{m=1}^{p} \beta_m s_{bi,t+j-m} + \epsilon^{\omega}_{\mathbf{bi,t+1}} & j = 1 \\ \alpha + \sum_{m=1}^{j-1} \beta_m \hat{\mathbf{s}}^{\omega}_{bi,t+j-m} + \sum_{m=j}^{p} \beta_m s_{bi,t+j-m} & 1 < j \leq p \\ \alpha + \sum_{m=1}^{p} \beta_m \hat{\mathbf{s}}^{\omega}_{bi,t+j-m} & j > p \end{cases} \tag{3.15}$$

where the error terms are assumed to be independent and normally distributed with mean zero and variance σ^2. Figure 3.15 shows the prediction of biomass for year 2015. It is observed that if we do not consider the prediction model then we will have to generate

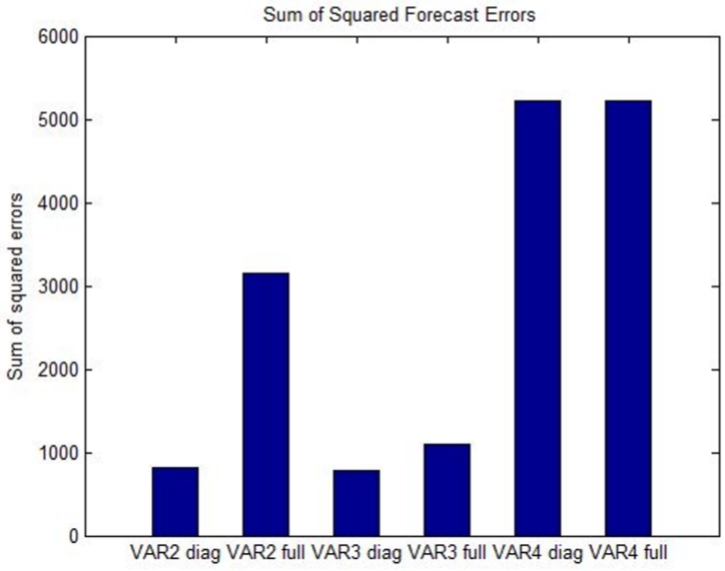


Figure 3.1

Sum of squared error

a large scenario set between $\overline{\epsilon}_t^1$ and $\underline{\epsilon}_t^1$ where $\hat{\epsilon}_t^1$ is the average biomass supply scenario for biomass type $b \in \mathcal{B}$ at time $t \in \mathcal{T}$. However, by developing a reliable prediction model we can generate a relatively small scenario set between $\overline{\epsilon}_t^2$ and $\underline{\epsilon}_t^2$ which is a subset of scenarios generated by $[\overline{\epsilon}_t^1, \underline{\epsilon}_t^1]$. We then use the prediction errors of biomass supply fluctuations to generate a tree of $|\mathcal{I}|$ scenarios for period time $t \in \mathcal{T}$ which is illustrated in Figure 3.16. We further use the scenario tree of the prediction error term (shown in Figure 3.16) and combine the forecasting and scenario generation to generate biomass supply scenarios for model **[CSC]**, as illustrated in Figure 3.17. We use the same approach to generate scenarios for both corn-stover and woody biomass.

3.2.2 Sample Average Approximation

The number of scenarios generated in **Section 3.2.1** is still large and commercial solvers such as CPLEX can not solve them. This motivates us to use Sample average approximation (SAA) algorithm to reduce the size of the problem defined by (3.1)-(4.13). SAA was previously applied to solve large scale supply chain network flow related problems, such as [116], [102], [16], [103] and others. SAA provides high quality feasible solutions along with the statistical estimation of their optimality gap. We can refer the work by Kleywegt et al.[60] for the proof of convergence properties of SAA and the work by Norkin et al. [87], [86] and Mark et al. [71] for the evaluation of statistical performance of SAA. In SAA, a smaller set of scenarios are solved repeatedly instead of solving the original problem with large number of scenarios and the procedure is repeated until a pre-specified tolerance gap is achieved. The first step in SAA will be generating random samples with $N < |\Omega|$ realizations of uncertain parameter and then approximate the recourse function with the sample average function $\frac{1}{N}\sum_{n=1}^{N}\mathbb{Q}(\mathbf{Y}, n)$. The problem (3.1)-(4.13) is now approximated by the following SAA problem:

$$\underset{Y\in\mathbf{Y}}{Minimize}\ \left\{\hat{\mathbf{g}}(Y) := \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt} + \frac{1}{N}\sum_{n=1}^{N}\mathbb{Q}(\mathbf{Y}, n)\right)\right\} \tag{3.16}$$

As the sample size (value of N) increases the optimal solution of (4.15), $\tilde{Y}_N$, and the optimal value $\boldsymbol{v}_N$, converges with probability one to an optimal solution of the original problem (3.1)-(4.13) [60]. Assuming that the SAA problem is solved within an absolute

optimality gap $\delta \geq 0$, we can estimate the sample size N needed to guarantee an ϵ-optimal solution to the true problem with probability at least equal to $(1-\alpha)$ as:

$$N \quad \geq \quad \frac{3\sigma^2_{max}}{(\epsilon-\delta)^2}\left(|\mathcal{L}||\mathcal{J}||\mathcal{T}|(log2) - log\alpha\right) \tag{3.17}$$

where $\epsilon > \delta$, $\alpha \in (0,1)$ and σ^2_{max} is a maximal variance of certain function differences [60]. Sample size estimation using equation (4.16) is too conservative for practical applications. Thus, one can choose a sample size N as a trade-off between the solution quality obtained by solving (4.15) to the original problem (3.1)-(4.13) and the computational burden needed to solve it. Solutions of the SAA problems with a sample size N provides a statistical lower and upper bounds for the original problem and the process terminates when the gap between the estimators falls below a pre-specified threshold value. The steps involved in solving **[CSC]** problem using Sample average approximation (SAA) are given below:

1. The smaller scenario set N are drawn randomly from the scenario tree N'.

2. Generate M independent biomass supply sample scenarios of size Ni.e., $\{\mathbf{s}^1_m(\omega), \mathbf{s}^2_m(\omega), ..., \mathbf{s}^N_m(\omega)\}$, $\forall m = 1, ..., M$, where $\mathbf{s} = \{s_{bit\omega}, \forall b \in \mathcal{B}, i \in \mathcal{I}, t \in \mathcal{T}, \omega \in \Omega\}$ and solve the corresponding SAA:

$$\underset{Y \in \mathbf{Y}}{Minimize} \ \left\{\hat{\mathbf{g}}(Y) := \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt} + \frac{1}{N}\sum_{n=1}^{N}\mathbb{Q}(\mathbf{Y}, n)\right)\right\} \tag{3.18}$$

The optimal objective value is denoted by $\mathbf{v}_N^m$ and the optimal solution by $\hat{\mathbf{Y}}_N^m$; $m = 1, ..., M$.

3. Compute the average of the optimal solutions obtained by solving all SAA problems, $\bar{\mathbf{v}}_M^N$ and variance, $\sigma^2_{\bar{\mathbf{v}}_M^N}$:

$$\begin{aligned} \bar{\mathbf{v}}_M^N &= \frac{1}{M}\sum_{m=1}^{M} \mathbf{v}_N^m \\ \sigma^2_{\bar{\mathbf{v}}_M^N} &= \frac{1}{(M-1)M}\sum_{m=1}^{M}\left(\mathbf{v}_N^m - \bar{\mathbf{v}}_M^N\right) \end{aligned}$$

where, $\bar{\mathbf{v}}_M^N$ provides a statistical lower bound on the optimal objective function value for the original problem (3.1)-(4.13) [87].

4. Pick a feasible first-stage solution $\tilde{Y} \in \mathbf{Y}$ obtained from **Step 1** of the SAA algorithm, e.g., one of the solutions from $\hat{\mathbf{Y}}_N^m$ and estimate the objective function value of the original problem **[CSC]** using a reference sample N' as follows:

$$\tilde{\mathbf{g}}_{N'}(\tilde{Y}) := \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}\tilde{Y}_{ljt} + \frac{1}{N'}\sum_{n=1}^{N'}\mathbb{Q}(\mathbf{Y}, n)\right) \tag{3.19}$$

The estimator $\tilde{\mathbf{g}}_{N'}(\tilde{Y})$ serves as an upper bound for the optimal objective function value of problem **[CSC]** which will be updated in each iteration if the value obtained is less than the value of the previous iteration. N' is another set of samples generated independently for the biomass supply scenarios, i.e., $\{\mathbf{s}^1(\omega), \mathbf{s}^2(\omega), ..., \mathbf{s}^{N'}(\omega)\}$, $\forall n = 1, ..., N'$. Typically, sample size N' is chosen much larger than the sample size

N in the SAA problems i.e., $N' \gg N$. We can estimate the variance of $\tilde{\mathbf{g}}_{N'}(\tilde{Y})$ as follows:

$$\sigma^2_{N'}(\tilde{Y}) \;=\; \frac{1}{(N'-1)N'} \sum_{n=1}^{N'} \Big\{ \sum_{t\in\mathcal{T}} \Big(\sum_{l\in\mathcal{L}} \sum_{j\in\mathcal{J}} \psi_{ljt} \tilde{Y}_{ljt} + \mathbb{Q}(\mathbf{Y}, n) \Big) - \tilde{\mathbf{g}}_{N'}(\tilde{Y}) \Big\}^2$$

5. Compute the optimality gap ($gap_{N,M,N'}(\tilde{Y})$) and its variance (σ^2_{gap}) using the estimators calculated in **Steps 2** and **3**.

$$\begin{aligned} gap_{N,M,N'}(\tilde{Y}) &= \tilde{\mathbf{g}}_{N'}(\tilde{Y}) - \bar{\mathbf{v}}^N_M \\ \sigma^2_{gap} &= \sigma^2_{N'}(\tilde{Y}) + \sigma^2_{\bar{\mathbf{v}}^N_M} \end{aligned}$$

The confidence interval for the optimality gap is then calculated as follows:

$$\tilde{\mathbf{g}}_{N'}(\tilde{Y}) - \bar{\mathbf{v}}^N_M + z_\alpha \Big\{ \sigma^2_{N'}(\tilde{Y}) + \sigma^2_{\bar{\mathbf{v}}^N_M} \Big\}^{1/2}$$

with z_α:= $\Phi^{-1}(1-\alpha)$, where $\Phi(z)$ is the cumulative distribution function of the standard normal distribution.

3.2.3 Progressive Hedging Algorithm

Solving **Step 1** in Sample average approximation algorithm is still considered challenging. This is because the step involves in solving a two-stage stochastic mixed integer linear programming model formulation with N scenarios. Even though the size of the SAA problem is significantly lower than the original problem **[CSC]**, it is still a challenging problem. To overcome this challenge, we solve each subproblems of the SAA problem

using a Progressive hedging algorithm (PHA) proposed by Rockafellar and Wets [99]. The PHA proceeds by applying a scenario decomposition technique based on the augmented Lagrangian relaxation scheme to solve a number of individual scenario subproblems and finally aggregating the individual scenario solutions. Progressive hedging algorithm has successfully applied to a number of different application areas, such as financial planning [84], fisheries management [47], surgery planning [44], hydrothermal operation planning [101], [14], and others. The readers are referred to the studies of Wallace and Helgason [119], Watson and Woodruff [123] for a detail discussion about the algorithm implementation.

Note that constraints (4.10) and (4.11) link the first-stage decisions with the second-stage decision variables. These constraints will not allow problem (4.17) separable by scenarios. To apply scenario decomposition scheme we define a new decision variable: $\{Y_{ljtn}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T},n\in N} \in \{0,1\}$. By doing this, a copy of the first-stage decision variable is created for each scenario $n \in N$. Problem (4.17) can now be rewritten as follows:

$$\underset{\mathbf{Y,Z,X,P,H,U}}{Minimize}\ \frac{1}{N}\sum_{n=1}^{N}\sum_{t\in\mathcal{T}}\Bigg\{\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljtn}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn}+$$
$$\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}(c_{bikt}+\delta_{bit}+\gamma_{bit})X_{biktn}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}(c_{bijkt}+\delta_{bit}+\gamma_{bit})X_{bijktn}+$$
$$\sum_{k\in\mathcal{K}}p_{bkt}P_{bktn}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Bigg)+\sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}\Bigg\}$$

Subject to: (4.5)-(4.9), (4.12), (4.13), and

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijktn} \leq \sum_{l\in\mathcal{L}} C_{lj}Y_{ljtn} \qquad \forall j\in\mathcal{J}, t\in\mathcal{T}, n\in N \tag{3.20}$$

$$Z_{jktn} \leq \sum_{l\in\mathcal{L}} \left\lceil \frac{C_{lj}}{C^{cap}} \right\rceil Y_{ljtn} \qquad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, n\in N \tag{3.21}$$

$$\sum_{l\in\mathcal{L}} Y_{ljtn} \leq 1 \qquad \forall j\in\mathcal{J}, t\in\mathcal{T}, n\in N \tag{3.22}$$

$$Y_{ljtn} = Y_{ljtm} \qquad \forall n,m\in N, n\neq m \tag{3.23}$$

$$Y_{ljtn} \in \mathbb{B} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T}, n\in N \tag{3.24}$$

Constraints (4.22) are referred to as *nonanticipativity* constraints which link the first- and second-stage decision variables and force all the scenarios to yield a same first-stage decision variables. This makes the model not separable to scenarios. To make the model separable and to apply Lagrangian relaxation, we need to rewrite the *nonanticipativity* constraints. Let $\{\bar{Y}_{ljt}\}_{\forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T}} \in \{0,1\}$ be the "overall design vector". The following constraints are equivalent to (4.22):

$$Y_{ljtn} = \bar{Y}_{ljt} \qquad \forall l\in\mathcal{L}, \forall j\in\mathcal{J}, t\in\mathcal{T}, n\in N \tag{3.25}$$

$$\bar{Y}_{ljt} \in \mathbb{B} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T} \tag{3.26}$$

Prior studies such as Wallace et al. [119], Mulvey et al. [84], and Haugen et al. [46] showed that constraints (25) and (26) are essentially equivalent to constraint (23).

We will now follow the decomposition scheme proposed by Rockafellar and Wets [99] to relax constraints (4.24) using an augmented Lagrangian strategy and obtain the following objective function:

$$\underset{\mathbf{Y,Z,X,P,H,U}}{Minimize}\ \frac{1}{N}\sum_{n=1}^{N}\sum_{t\in\mathcal{T}}\Bigg\{\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljtn}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn}+$$

$$\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}(c_{bikt}+\delta_{bit}+\gamma_{bit})X_{biktn}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}(c_{bijkt}+\delta_{bit}+\gamma_{bit})X_{bijktn}+$$

$$\sum_{k\in\mathcal{K}}p_{bkt}P_{bktn}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Bigg)+\sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}+\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\lambda}_{ljtn}(Y_{ljtn}-\bar{Y}_{ljt})+$$

$$\frac{1}{2}\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\phi}(Y_{ljtn}-\bar{Y}_{ljt})^2\Bigg\}$$

where $\{\boldsymbol{\lambda}_{ljtn}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T},n\in N}$ defines the Lagrangian multipliers for the relaxed constraints and $\boldsymbol{\phi}$ defines a penalty ratio. Given the binary requirements of variables $\{Y_{ljtn}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T},n\in N}$ and $\{\bar{Y}_{ljt}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ the quadratic term $\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\phi}(Y_{ljtn}-\bar{Y}_{ljt})^2$ shown in the above objective function can be reduced as follows:

$$\begin{aligned}\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\phi}(Y_{ljtn}-\bar{Y}_{ljt})^2 &= \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\Big(\boldsymbol{\phi}(Y_{ljtn})^2-2\boldsymbol{\phi}Y_{ljtn}\bar{Y}_{ljt}+\boldsymbol{\phi}(\bar{Y}_{ljt})^2\Big)\\ &= \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\Big(\boldsymbol{\phi}Y_{ljtn}-2\boldsymbol{\phi}Y_{ljtn}\bar{Y}_{ljt}+\boldsymbol{\phi}\bar{Y}_{ljt}\Big)\end{aligned}$$

The objective function therefore can be reduced as follows:

$$\underset{\mathbf{Y,Z,X,P,H,U}}{Minimize}\ \frac{1}{N}\sum_{n=1}^{N}\sum_{t\in\mathcal{T}}\Bigg\{\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}(\psi_{ljt}+\boldsymbol{\lambda}_{ljtn}-\boldsymbol{\phi}\bar{Y}_{ljt}+\frac{\boldsymbol{\phi}}{2})Y_{ljtn}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn}+$$

$$\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}(c_{bikt}+\delta_{bit}+\gamma_{bit})X_{biktn}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}(c_{bijkt}+\delta_{bit}+\gamma_{bit})X_{bijktn}+$$

$$\sum_{k\in\mathcal{K}}p_{bkt}P_{bktn}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Bigg)+\sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}-\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\lambda}_{ljtn}\bar{Y}_{ljt}-\frac{1}{2}\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\phi}\bar{Y}_{ljt}\Bigg\}$$

When the value of the overall plan $\{\bar{Y}_{ljt}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ is fixed, the last two terms of the objective function (4.26) becomes constant and thus can be decomposed into a series of

deterministic problem for each scenario $n \in N$. The overall problem can be formulated as follows:

$$\textbf{[CSC(PHA)]}\ \underset{\mathbf{Y,Z,X,P,H,U}}{Minimize}\ \sum_{t\in\mathcal{T}}\Bigg\{\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}(\psi_{ljt}+\boldsymbol{\lambda}_{ljtn}-\boldsymbol{\phi}\bar{Y}_{ljt}+\frac{\boldsymbol{\phi}}{2})Y_{ljtn}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn}+$$

$$\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}(c_{bikt}+\delta_{bit}+\gamma_{bit})X_{biktn}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}(c_{bijkt}+\delta_{bit}+\gamma_{bit})X_{bijktn}+$$

$$\sum_{k\in\mathcal{K}}p_{bkt}P_{bktn}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Bigg)+\sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}\Bigg\}$$

Subject to

$$\sum_{k\in\mathcal{K}}X_{biktn}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}X_{bijktn} \leq s_{bitn} \qquad \forall b\in\mathcal{B}, i\in\mathcal{I}, t\in\mathcal{T} \quad (3.27)$$

$$\sum_{i\in\mathcal{I}}X_{biktn}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}X_{bijktn}+$$

$$(1-\alpha_b)H_{b,k,t-1,n} = H_{bktn}+P_{bktn} \qquad \forall b\in\mathcal{B}, k\in\mathcal{K}, t\in\mathcal{T} \quad (3.28)$$

$$\sum_{b\in\mathcal{B}}P_{bktn}+U_{ktn} = d_{kt} \qquad \forall k\in\mathcal{K}, t\in\mathcal{T} \quad (3.29)$$

$$\sum_{b\in\mathcal{B}}H_{bktn} \leq h_k^{cap} \qquad \forall k\in\mathcal{K}, t\in\mathcal{T} \quad (3.30)$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}X_{bijktn} \leq c^{cap}Z_{jktn} \qquad \forall (j,k)\in\mathcal{A}_2, t\in\mathcal{T} \quad (3.31)$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijktn} \leq \sum_{l\in\mathcal{L}}C_{lj}Y_{ljtn} \qquad \forall j\in\mathcal{J}, t\in\mathcal{T} \quad (3.32)$$

$$Z_{jktn} \leq \sum_{l\in\mathcal{L}}\left\lceil\frac{C_{lj}}{c^{cap}}\right\rceil Y_{ljtn} \qquad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T} \quad (3.33)$$

$$\sum_{l\in\mathcal{L}}Y_{ljtn} \leq 1 \qquad \forall j\in\mathcal{J}, t\in\mathcal{T} \quad (3.34)$$

$$Y_{ljtn} \in \mathbb{B} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T} \quad (3.35)$$

$$Z_{jktn} \in \mathbb{Z}^+ \qquad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T} \quad (3.36)$$

$$X_{bijktn}, X_{biktn}, H_{bktn}, P_{bktn}, U_{ktn} \in \mathbb{R}^+ \quad \forall b\in\mathcal{B}, i\in\mathcal{I}, j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T} \quad (3.37)$$

Let r denote the index for the iteration number for the progressive hedging algorithm and $\{\boldsymbol{\lambda}^r_{ljtn}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T},n\in N}$ is the lagrangian multipliers and $\boldsymbol{\phi}^r$ is the penalty parameter at iteration r. The general idea of the basic Progressive hedging algorithm is to solve N deterministic **[CSC(PHA)]** problem and obtain the consensus parameter $\{\bar{Y}^r_{ljt}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$. If the gap between the binary variable Y^r_{ljtn} and the consensus parameter $\bar{Y}^r_{ljt}$ falls below a threshold value ϵ (i.e., $\epsilon = 0.001$) for each $l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$ then the algorithm terminates; otherwise, the values of $\boldsymbol{\lambda}^r_{ljtn}$ and $\boldsymbol{\phi}^r$ are updated using equations (3.38) and (4.39) and the process continues.

$$\boldsymbol{\lambda}^r_{ljtn} \leftarrow \boldsymbol{\lambda}^{r-1}_{ljtn} + \boldsymbol{\phi}^{r-1}_{ljt}(Y^r_{ljtn} - \bar{Y}^{r-1}_{ljt}) \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T} \quad (3.38)$$

$$\boldsymbol{\phi}^r \leftarrow \alpha\boldsymbol{\phi}^{r-1} \quad (3.39)$$

where $\boldsymbol{\lambda}^0_{ljtn}$ is set to zero for each scenario $n \in N$; $\boldsymbol{\phi}^0$ is set to a fixed positive value which ensures that $\boldsymbol{\phi}^r \to \infty$ as the number of iteration r increases and the constant parameter $\alpha > 1$. A pseudo-code of the basic Progressive hedging algorithm is provided in **Algorithm 1**.

Termination Criteria: The Progressive hedging algorithm terminates when one of the following condition is satisfied. These criteria are obtained from the studies of Watson and Woodruff [123], Crainic et al.[19], and Chen and Fan [17]:

- $\frac{1}{N}\sum_{n=1}^{N}\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}|Y^r_{ljtn} - \bar{Y}^r_{ljt}| \leq \epsilon$; where ϵ is a pre-specified tolerance gap

Algorithm 1: Progressive Hedging Algorithm

Initialize, $r \leftarrow 1$, ϵ, $\{\boldsymbol{\lambda}^r_{ljtn}\}_{\forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}, n \in N} \leftarrow 0$, $\boldsymbol{\phi}^r \leftarrow \boldsymbol{\phi}^0$
terminate $\leftarrow$ **false**
while (*terminate* $=$ **false**) **do**
    for $n = 1$ to N
        Solve **[CSC(PHA)]** and obtain $\{Y^r_{ljtn}\}_{\forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}, n \in N}$
    end for
    Calculate the consensus parameter:
        $\bar{Y}^r_{ljt} \leftarrow \frac{1}{N} \sum_{n=1}^{N} Y^r_{ljtn}; \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$
    if $(r > 1)$ **then**
        Update the lagrangian parameter:
            $\boldsymbol{\lambda}^r_{ljtn} \leftarrow \boldsymbol{\lambda}^{r-1}_{ljtn} + \boldsymbol{\phi}^{r-1}_{ljt}(Y^r_{ljtn} - \bar{Y}^{r-1}_{ljt}); \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$
        Update the penalty parameter:
            $\boldsymbol{\phi}^r \leftarrow \alpha \boldsymbol{\phi}^{r-1}$ and $\alpha > 1$
    end if
    if $((Y^r_{ljtn} - \bar{Y}^{r-1}_{ljt})_{\forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}} \leq \epsilon)$ **then**
        terminate $\leftarrow$ **true**
    end if
    $r \leftarrow r + 1$
end while

- 10 consecutive non-improvement iterations
- Maximum iteration limit is reached (i.e., $iter^{max}$ = 100)
- Maximum time limit is reached (i.e., $time^{max}$ = 10,800 CPU seconds)

3.2.4 Enhanced Progressive Hedging Algorithm

Our initial computational experimentation with Progressive hedging algorithm with relatively large network size problem exposes its inability to converge within a reasonable amount of time. This motivates us to explore additional enhancement techniques to improve the convergence and stability of the Progressive hedging algorithm. Therefore, the following subsections will present some enhancement techniques to solve problem **[CSC(PHA)]** efficiently.

3.2.4.1 Penalty Parameter Updating

In this study, we set a constant value for the penalty parameter ϕ. Prior research indicates that setting of ϕ strongly influences the value of the objective function [17, 50]. It is observed that when the value of ϕ is too large then the algorithm convergence fast to a suboptimal solution. On the other hand, if the value of ϕ is too low then the algorithm convergence slowly to a near optimal solution. To tackle this issue, we used a method proposed by Hvattum and Lokketangen [53] to dynamically adjust the value of ϕ over iterations based on comparing the convergence rate of the algorithm at iterations r and $r-1$. Let $\boldsymbol{\Delta}_1^r$ and $\boldsymbol{\Delta}_2^r$ be indicators of the convergence rates in the dual space and in the primal space, respectively. Thus, the penalty value can be updated as follows:

$$\boldsymbol{\Delta}_1^r = \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}(Y_{ljt}^r - \bar{Y}_{ljt}^r)^2 \tag{3.40}$$

$$\boldsymbol{\Delta}_2^r = \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}(\bar{Y}_{ljt}^r - \bar{Y}_{ljt}^{r-1})^2 \tag{3.41}$$

$$\boldsymbol{\phi}^r = \begin{cases} \varphi\boldsymbol{\phi}^{r-1} & \text{if } \boldsymbol{\Delta}_1^r - \boldsymbol{\Delta}_1^{r-1} > 0 \\ \frac{1}{\varphi}\boldsymbol{\phi}^{r-1} & \text{else if } \boldsymbol{\Delta}_2^r - \boldsymbol{\Delta}_2^{r-1} > 0 \end{cases} \tag{3.42}$$

where φ is a constant parameter which value is set to $\varphi > 1$.

3.2.4.2 Heuristic Strategies

We will now use few heuristic strategies by modifying the fixed cost of using multimodal facilities in solving problem **[CSC(PHA)]** at each iteration r of **Algorithm 1** (see

Crainic et al. [19] for details). The first strategy is termed as *global heuristics* since this strategy modifies the fixed cost of using a multi-modal facility at the end of each iteration. Remember that the problem **[CSC(PHA)]** is composed of a series of N deterministic subproblems. At the end of each iteration r in **Algorithm 1**, one obtain the values of the consensus parameter $\{\bar{Y}^r_{ljt}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ which provide an indication of how many times a multi-modal facility was used in the previous iterations. A higher value of $\bar{Y}^r_{ljt}$ translates the fact that the multi-modal facility j of capacity l at time period t was used in most of the previous iterations. Conversely, a lower value of $\bar{Y}^r_{ljt}$ indicates that the multi-modal facility j of capacity l at time period t was not a favorable decision in most of the previous iterations. Let $\overline{c}$ and $\underline{c}$ are the two parameters that define the upper and lower threshold value. Therefore, if the value of $\bar{Y}^r_{ljt}$ is greater that the threshold value $\overline{c}$, then lowering the fixed cost of using the multi-modal facility will attract the subproblems to use the facility in the coming iterations. Similarly, if the value of $\bar{Y}^r_{ljt}$ is lower that the threshold value $\underline{c}$, then increasing the fixed cost of using the multi-modal facility will avoid the subproblems use the facility in the coming iterations. This will fix the decisions of using few multi-modal facilities to either one or zero and thus will help to reduce the size of the problem. The adjustment strategy is shown below:

$$\psi^r_{ljt} = \begin{cases} \beta\psi^{r-1}_{ljt} & \text{if } \bar{Y}^{r-1}_{ljt} < \underline{c} \\ \frac{1}{\beta}\psi^{r-1}_{ljt} & \text{if } \bar{Y}^{r-1}_{ljt} > \overline{c} \\ \psi^{r-1}_{ljt} & \text{Otherwise} \end{cases} \tag{3.43}$$

where ψ^r_{ljt} represents the modified fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$ and at iteration r; $\underline{c}$ and $\overline{c}$ are the two constant parameters whose values are set to $0 < \underline{c} < 0.3$ and $0.7 < \overline{c} < 1$; and β is a constant parameter whose value is set to $\beta > 1$.

The above *global heuristics* strategy can be pushed even further to modify the fixed cost of using multi-modal facilities locally within the scenario level. Crainic et al. [19] named this strategy as *local heuristics* strategy since the modification of the fixed cost only impacts the subproblem of scenario n at current iteration r. This strategy emphasizes in modifying the fixed cost of using the multi-modal facility at scenario $n \in N$ in iteration r if the gap between Y^r_{ljtn} and $\bar{Y}^r_{ljt}$ is sufficiently large. The local adjustment strategy applied to **Algorithm 1** is shown below:

$$\psi^r_{ljtn} = \begin{cases} \beta\psi^r_{ljt} & \text{if } |Y^{r-1}_{ljtn} - \bar{Y}^r_{ljt}| \geq c^{far} \text{ and } Y^{r-1}_{ljtn} = 1 \\ \frac{1}{\beta}\psi^r_{ljt} & \text{if } |Y^{r-1}_{ljtn} - \bar{Y}^r_{ljt}| \geq c^{far} \text{ and } Y^{r-1}_{ljtn} = 0 \\ \psi^r_{ljt} & \text{Otherwise} \end{cases} \tag{3.44}$$

where ψ^r_{ljtn} represents the modified fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$ under scenario $n \in N$ and at iteration r; c^{far} is a threshold at which point a local adjustment to the fixed cost of using multi-modal facility is applied and is set to $0.5 < c^{far} < 1$; and β is a constant parameter whose value is set to $\beta > 1$.

3.2.4.3 Rolling Horizon (RHA) Heuristic

Note that, in Progressive hedging algorithm we need to solve problem **[CSC(PHA)]** N times which is a deterministic, multi-time period facility location problem. Solving this problem is still considered challenging since the problem is a special case of a capacitated facility location problem which is known to be an $\mathcal{NP}$-hard problem [69]. In this section we solve this problem using a heuristic approach proposed by Balasubramanian and Grossman [7] and Kostina et al.[62]. This approach decomposes problem **[CSC(PHA)]** into a series of small subproblems where each subproblem includes few consecutive time periods which are drawn from the overall planning horizon. The algorithm terminates when all the subproblems are investigated. The overall algorithm is shown in **Algorithm 2**.

Let t_0^r denote the starting time period of subproblem r. Let M^r denote the number of time periods comprised in subproblem r. We can set a fixed or variable size of M^r for each subproblem. Each approximate subproblem of the rolling horizon algorithm is denoted by **[CSC(PHA(r))]**. Now, for each scenario $n \in N$ the approximate subproblems are solved by setting the variables as: (i) $\{Y_{ljtn}\}_{\forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}} \in \{0,1\}$ and $\{Z_{jktn}\}_{\forall j \in \mathcal{J}, k \in \mathcal{K}, t \in \mathcal{T}} \in \mathbb{Z}^+$ for $t_0^r \leq t \leq t_0^r + M^r$, (ii) $0 \leq Y_{ljtn} \leq 1$ and $Z_{jktn} \in \mathbb{R}^+$ for $t > t_0^r + M^r$. Once a subproblem is solved, we fix the values of $Y_{ljtn}^r = Y_{ljtn}^{r-1}; \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$ and $Z_{jktn}^r = Z_{jktn}^{r-1}; \forall j \in \mathcal{J}, k \in \mathcal{K}, t \in \mathcal{T}$ for $t < t_0^r$ and update the step size r. The process terminates when all the subproblems are solved. **Figure 3.2** shows an example of using the rolling horizon approach to solve a three time period problem.

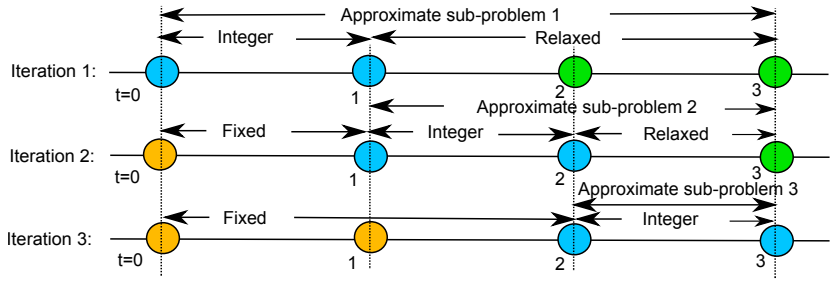


Figure 3.2

Application of a rolling horizon strategy for a three time period problem

3.3 Computational Study and Managerial Insights

In order to test the performance of the algorithms proposed in **Section 3.2** we develop a case study where we use Mississippi and Alabama as a testing ground for the analysis. All the algorithms are coded in GAMS 24.2.1 [39] and executed on a desktop computer with Intel Core i7 3.50 GHz processor and 32.0 GB RAM. The optimization solver used is ILOG CPLEX 12.6. In this section, we first provide details about the data used to develop the case study. Next, we conduct a computational study on model **[CSC]** to test the performance of the algorithms and finally we present results from the case study and draw managerial insights.

3.3.1 Data Description

***Biomass Supply and Demand*:** Two states in the Southern United States Mississippi (MS) and Alabama (AL) are considered as a testing ground for this case study. The two main biomass feedstocks available in these two states are corn stover and forest residues. Biomass are not available all year around. Corn stover are available only in fall season (from September until November) whereas woody biomass are available all year around

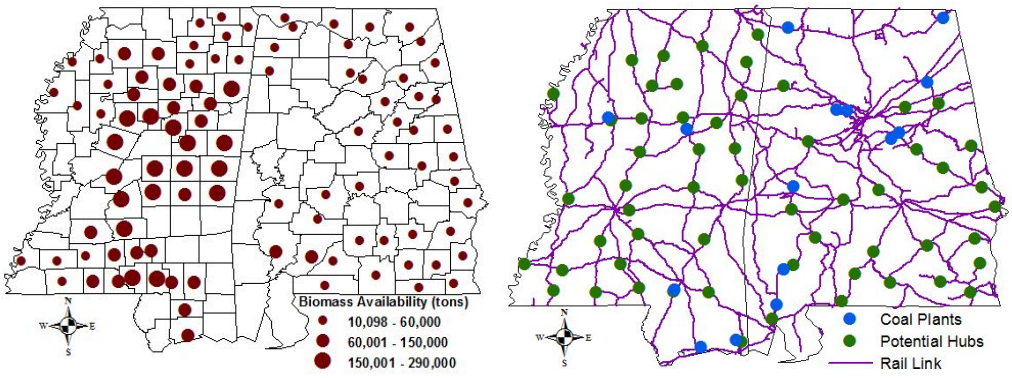
Algorithm 2: Rolling Horizon (RHA) Heuristic

$t_0^r = 0$, $r \leftarrow 1$, M^r, *terminate* $\leftarrow$ **false**
while (*terminate* $=$ **false**) **do**
  Set:
    $Y_{ljtn} \in \{0, 1\}$ and $Z_{jktn} \in \mathbb{Z}^+$ for $t_0^r \leq t \leq t_0^r + M^r$
    $0 \leq Y_{ljtn} \leq 1$ and $Z_{jktn} \in \mathbb{R}^+$ for $t > t_0^r + M^r$
  Solve the approximate sub-problem **[CSC(PHA)]** using CPLEX
  if($t_0 > |\mathcal{T}|$) **then**
    stop $\leftarrow$ **true**
  else
    Fixing the values of Y_{ljtn} and Z_{jktn} for $t < t_0^r$
  end if
  $r \leftarrow r + 1$
end while

except three months during the winter (December to February). The historical availability data of corn stover and woody biomass is reported by the Knowledge Discovery Framework (KDF) database of United States Department of Energy [10]. The data was further processed at Idaho National Lab (INL) to calculate the amount of densified biomass available in Mississippi and Alabama. Figure 4.5a shows the distribution of densified biomass available in this region where the position and size of the dot represents the location and amount of biomass available in each supply site $i \in \mathcal{I}$. In total 99 suppliers spread throughout Mississippi and Alabama and the counties that produce more than 10,000 tons of biomass each year are only considered here.

The data about coal-fired power plant locations and capacities is obtained from the National Energy Technology Laboratory [110]. We have considered in total 15 coal plants which are distributed in Mississippi and Alabama and their geographic distribution are shown in Figure 4.5b. We assume that power plants will replace 6% of coal's demand with densified biomass to produce renewable electricity. This results in a demand of 5.6 million

tons per year (MTY) of densified biomass for Mississippi and Alabama. This amount is estimated based on the nameplate capacity, efficiency and operating hours of the plant, and the lower heating values of coal and biomass [57].



(a) Biomass distribution (b) Network representation

Figure 3.3

Biomass distribution and facility locations

***Investment Costs*:** This study considers a total of 57 potential locations for rail ramps whose geographical location and distribution in the states of Mississippi and Alabama are shown in Figure 4.5b. The annualized fixed cost for a rail ramp of capacity 1.05 MTY that sends a single railcar is set to $54,949/year [70]. We considered four different rail ramp capacities: l = 0.5 MTY, 0.75 MTY, 1.0 MTY, and 1.25 MTY. These costs are estimated based on a lifetime of 30 years, and a discount factor of 10% is assumed.

***Transportation Costs*:** This study assumes that trucks are used to transport biomass from a feedstock supplier $i \in \mathcal{I}$ to a multi-modal facility $j \in \mathcal{J}$. Trucks can further be used to transport biomass directly from a feedstock supplier $i \in \mathcal{I}$ to a coal plant $k \in \mathcal{K}$. Table

4.2 shows the major cost components used to calculate the unit cost of truck transportation which are obtained from the study of Parker et al. [88].

Table 3.1

Truck transportation cost components [88]

Item	Value	Unit
Loading/unloading	5.0	$/wet ton
Time dependent	29.0	$/hr/truckload
Distance dependent	1.20	$/mile/truckload
Truck capacity	25	wet tons/truckload
Average travel speed	40	miles/hour

This study further assumes that railcar will be used to transport biomass from a multi-modal facility $j \in \mathcal{J}$ to a coal plant $k \in \mathcal{K}$. We used \$2,248 (ξ_{jkt}) as the fixed cost and \$1.12 as the unit transportation cost per mile traveled for a single railcar of capacity 100 tons [42]. Finally, we used Arc GIS Desktop 10 to create a transportation network and used this network to identify shortest paths between the sources and destinations. The network includes actual railways as well as local, rural, urban roads and major highways in Mississippi and Alabama.

3.3.2 Analyzing the Performance of Solution Algorithms

This section presents our computational experience in solving model **[CSC]** using the algorithms proposed in Section 3.2. To help the readers follow our approaches, we have used the following notations to represent the algorithms:

- **[SAA]**: Sample Average Approximation (SAA) algorithm (described in Section 4.3.2)

- **[SAA+PHA]**: Sample average approximation algorithm where the subproblem of the **[SAA]** is solved using Progressive Hedging algorithm (PHA) (described in Section 3.2.3)
- **[SAA+PHA+HEU]**: Sample average approximation algorithm where the subproblem of the **[SAA]** is solved using an enhanced Progressive Hedging algorithm (PHA) (enhancement techniques described in Section 4.3.4.1 and 3.2.4.2)
- **[SAA+PHA+HEU+RHA]**: Sample average approximation algorithm where the subproblem of the **[SAA]** is solved using an enhanced Progressive Hedging algorithm (PHA) (enhancement techniques described in Section 4.3.4.1, 3.2.4.2, and 3.2.4.3)

The algorithms presented above are terminated when at least one of the following condition is met: (a) the optimality gap (i.e., $\epsilon = |UB - LB|/UB$) falls below a threshold value ϵ = 0.01; or (b) the maximum time limit $time^{max}$ = 10,800 (in CPU seconds) is reached; or (c) the maximum number of iteration $iter^{max}$ = 100 is reached. To terminate the Progressive Heading algorithm, we have used some additional stopping criteria which are described at the end of Section 3.2.3. The sizes of the deterministic equivalent problems of model **[CSC]** are presented in Table 4.3.

Table 3.2

Problem size of the test instances

Case	$\|\mathcal{I}\|$	$\|\mathcal{J}\|$	$\|\mathcal{K}\|$	$\|\mathcal{L}\|$	$\|\mathcal{B}\|$	$\|\mathcal{T}\|$	Binary Variables	Integer Variables	Continuous Variables	Total Variables	No. of Constraints
1	99	57	15	4	2	4	912	3,420	689,340	693,672	8,329
2	99	57	15	4	2	8	1,824	6,840	1,378,680	1,387,344	16,657
3	99	57	15	4	2	12	2,736	10,260	2,068,020	2,081,016	24,985

The first set of experiments (reported in Table 3.3) show how using different enhancement techniques speed up the convergence and improves the quality of the Progressive Hedging algorithm (**[PHA]**). The enhancement techniques used are: (i) **[PHA+HEU]** that incorporates heuristics strategies (described in Section 3.2.4.2) and penalty parameter updating techniques (described in Section 4.3.4.1) in **[PHA]** algorithm and

(ii) **[PHA+HEU+RHA]** that incorporates rolling horizon algorithm (described in section 3.2.4.3) along with heuristics strategies (described in Section 3.2.4.2), penalty parameter updating techniques (described in Section 4.3.4.1) in **[PHA]** algorithm. We consider three scenario sizes $N = \{20, 30, 40\}$ and three time periods $|\mathcal{T}| = \{4, 8, 12\}$ to obtain 9 different problem instances. Additionally, we have tested the performance of the enhancement techniques under both normal and uniform distribution to obtain an overall 18 different problem instances. We do not present the results obtained from CPLEX. This is because CPLEX gets *out of memory* in solving all the problem instances described in Table 3.3. In all of our experimental results, if the algorithms are solved in less than the stopping criteria ϵ then we highlighted the algorithm which gave the smallest running time. Otherwise, if such a quality solution is not found within the maximum time or iteration limit then the algorithm with the smallest optimality gap is highlighted. The results indicate that implementing the cuts presented in section 3.2.4 substantially improves the performance of the Progressive hedging algorithm. It is observed that algorithm **[PHA+HEU]** solves 13 out of 18 problem instances in 1% optimality gap within the specified time limit. On the other hand, the standard Progressive hedging algorithm **[PHA]** fails to solve any problem instances within the pre-specified optimality gap. We observe that the perfor-

mance can be enhanced even further by solving algorithm **[PHA+HEU]** in a rolling horizon framework (**[PHA+HEU+RHA]**). Algorithm **[PHA+HEU+RHA]** is capable of solving 17 out of 18 problem instances in less that 1% optimality gap within the specified time limit. On average, algorithm **[PHA+HEU+RHA]** is 1.42 times faster than algorithm **[PHA+HEU]**. Furthermore, algorithm **[PHA+HEU+RHA]** drops the average optimality gap to 0.82% from 2.00% provided by the **[PHA+HEU]** algorithm. Note that algorithm **[PHA+HEU+RHA]** does not guarantee to generate valid lower bound for algorithm **[PHA]**. Therefore, in Table 3.3 we use the lower bound of **[PHA+HEU]** algorithm to present an optimality gap for **[PHA+HEU+RHA]** algorithm i.e., $100*(UB_{\textbf{[PHA+HEU+RHA]}} - LB_{\textbf{[PHA+HEU]}})/UB_{\textbf{[PHA+HEU+RHA]}}\%$. Table 3.3 shows that even though algorithm **[PHA+HEU+RHA]** terminates with an ϵ-optimal solution, the quality of solution produced by the **[PHA+HEU+RHA]** algorithm is consistently high.

We now show the benefits of using different accelerated techniques in each replication of the sample average approximation **[SAA]** algorithm. Figure 3.4 shows the average computing time spent in each replication to solve the instances with a problem size of $|\mathcal{I}| = |\mathcal{J}| = |\mathcal{K}| = 10$, $|\mathcal{L}| = |\mathcal{T}| = 4$, $|\mathcal{B}| = 2$, $N = 10$, $M = 50$ and the scenarios are generated using a normal distribution. It is observed that the computational time is significantly reduced by using the **[SAA+PHA+HEU]** algorithm over algorithm **[SAA+PHA]**. The computational time can be reduced even further by incorporating the rolling horizon framework in the **[SAA+PHA+HEU]** algorithm. On average, algorithm **[SAA+PHA+HEU+RHA]** is 1.4 and 1.2 times faster than the **[SAA+PHA+HEU]** and **[SAA+PHA]** algorithms, respectively.

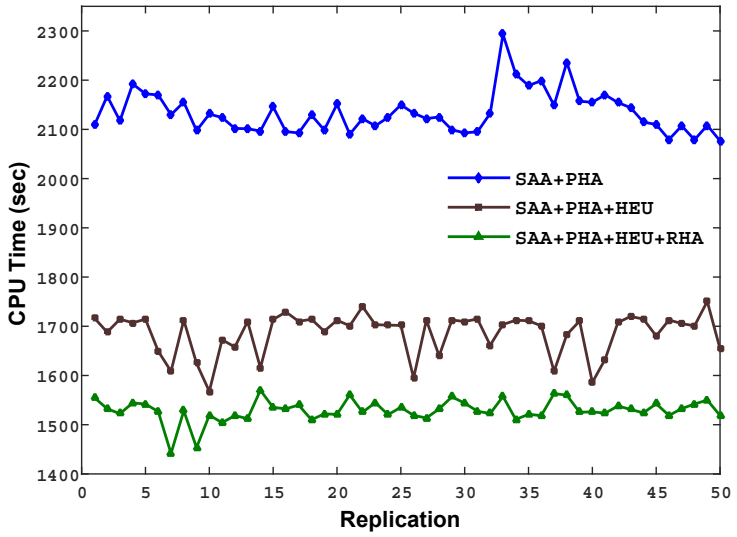


Figure 3.4

Comparison of solution time in each replication of the SAA algorithm

Table 3.4 presents the results from solving model **[CSC]** using the algorithms proposed in Section 3.2. The problems were solved using *Case 3* problem instances obtained from Table 4.3 with varying sample size N and replication number M in the **[SAA]** algorithm. The size of the large scenario set was chosen equal to $N' = 500$ to evaluate the SAA gap. We do not present results obtained from CPLEX since CPLEX gets *out of memory* in solving all the problem instances reported in Table 3.4. Results indicate that, **[SAA]** is able to solve only 1 out of 12 problem instances in less than 1.0% optimality gap within the specified time limit. The performance can be improved slightly by solving the subproblems of the **[SAA]** algorithm using **[PHA]** algorithm. It is observed that algorithm **[SAA+PHA]** is now able to solve 3 out of 12 problem instances in less than 1.0% optimality gap within the specified time limit. The benefits of using the algorithms become more evident when

the enhancement techniques (developed in section 3.2.4) implemented in **[PHA]** algorithm is used to solve the subproblems of the **[SAA]** algorithm. The overall average optimality gap for the the **[SAA+PHA+HEU]** algorithm is reported as 2.21%, with 7 out of 12 problem instances is solved in less than 1.0% optimality gap within the specified time limit. On the other hand, the overall average optimality gap for the **[SAA+PHA+HEU+RHA]** algorithm is reported as 1.03%, with 9 out of 12 problem instances is solved in less than 1.0% optimality gap within the specified time limit. It is important to note that algorithm **[SAA+PHA+HEU+RHA]** saves 11.7% time over algorithm **[SAA+PHA+HEU]** in producing the optimality gaps. In overall, the **[SAA+PHA+HEU+RHA]** algorithm seems to offer high quality solutions consistently within the experimental range.

3.3.3 Experimental Results

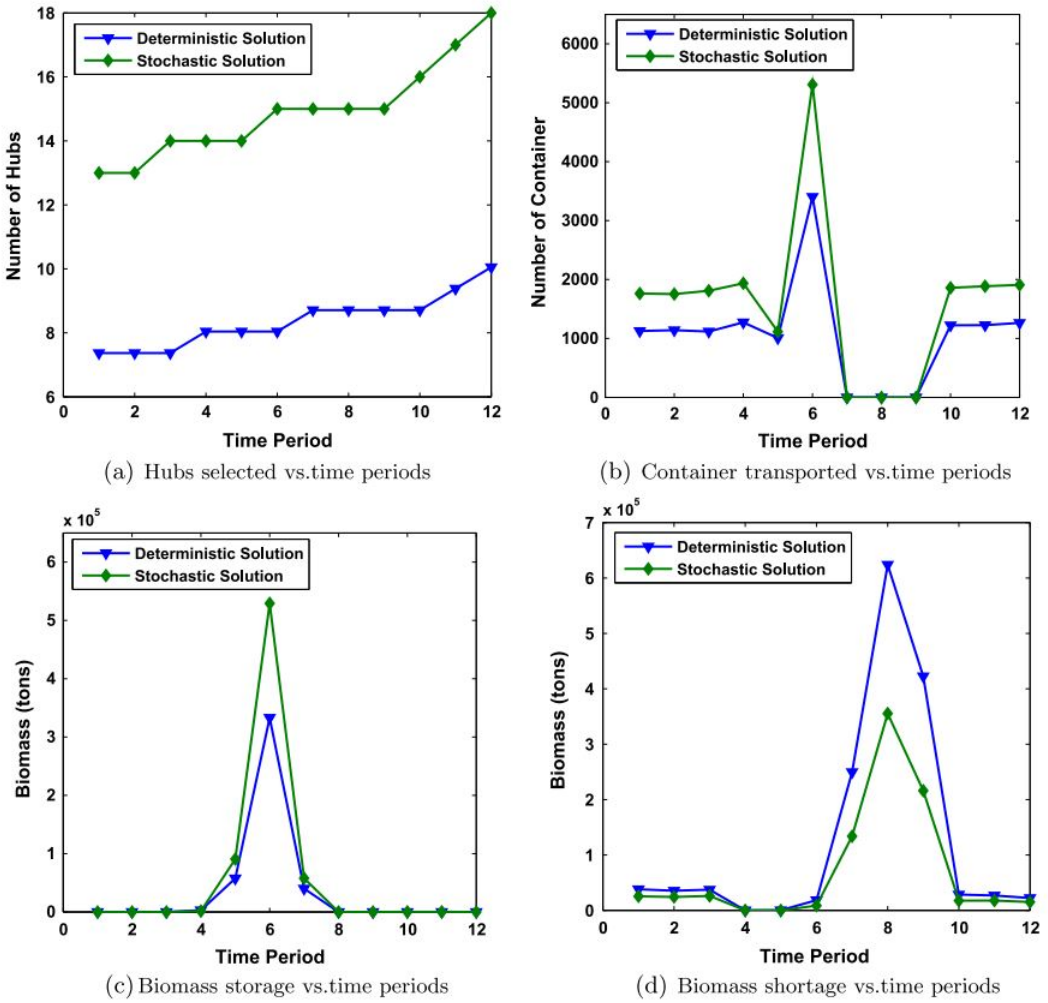
3.3.3.1 Performance evaluation of stochastic and deterministic solution:

We first justify the use of stochastic programming approach in solving model **[CSC]** over its deterministic equivalent solution. We use the average biomass supply $\overline{s}_{bit} = s_{bit\omega}/|\Omega|$ where $s_{bit\omega}$ is obtained from the stochastic programming approach to generate the deterministic equivalent solutions. Figure 3.5 shows the key supply chain network decisions with and without considering biomass supply uncertainty. It is observed that the stochastic programming approach selects on average 3.65% additional multi-modal facilities in each time period $t \in \mathcal{T}$ of the year to cope against feedstock supply variability. Figure 3.6 shows the comparison of network configuration between stochastic and deterministic approaches. This results an additional 2.82% (on average) containers transported between the multi-modal facilities and coal plants in each time period $t \in \mathcal{T}$ of the year.

Furthermore, the coal plants store on average 9.05% additional biomass between months of December to February (t=5-7) when neither corn-stover nor woody biomass is available. This results in an overall 44.19% decrease in biomass shortage quantity and thus reduces the unit delivery cost of biomass to \$40.97/ton (stochastic programming solution) from \$45.11/ton (deterministic equivalent solution). In summary, it is observed that the biomass supply uncertainty highly impacts the supply chain network decisions.

3.3.3.2 Impact of biomass supply variation levels on system performance:

The second set of experiments shows how different levels of supply variation impact the biomass supply chain network performance. To achieve this goal, we create three different realistic scenarios. In the first scenario (base case), we solve model **[CSC]** using the input parameters discussed in Section 4.4.1. The second (high variations) and third scenarios (low variations) are created by setting ϵ = 50% for *high biomass supply variation* levels and ϵ = 5% for *low biomass supply variation* levels and the scenarios are generated using a uniform distribution between $[\overline{s}_{bit}(1-\epsilon), \overline{s}_{bit}(1+\epsilon)]$ for each biomass type $b \in \mathcal{B}$ in location $i \in \mathcal{I}$ and in time period $t \in \mathcal{T}$. Note that $\overline{s}_{bit}$ represents the mean biomass supply scenario for each biomass type $b \in \mathcal{B}$ in location $i \in \mathcal{I}$ and in time period $t \in \mathcal{T}$. Figure 3.7 presents the impact of different biomass supply variation levels on supply chain network performance. Results indicate that as the level of biomass supply variability increases the number of multi-modal facilities ($|\mathbf{Y}|$) and the number of containers ($|\mathbf{Z}|$) transporting between the links $(j,k) \in \mathcal{A}_2$ in each time period $t \in \mathcal{T}$ also increases. Furthermore, we observe that to cope against high biomass supply variability levels, the coal plants



(a) Hubs selected vs.time periods

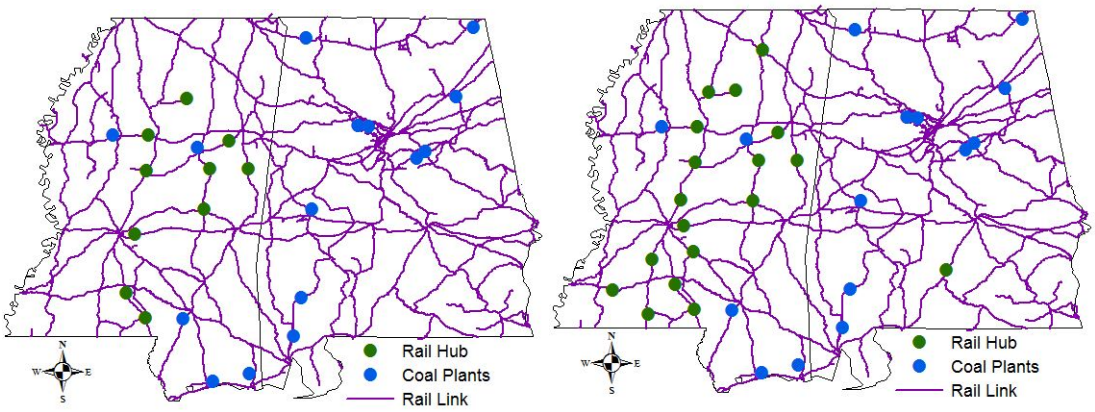
(b) Container transported vs.time periods

(c) Biomass storage vs.time periods

(d) Biomass shortage vs.time periods

Figure 3.5

Key supply chain network decisions with and without considering biomass supply uncertainty



(a) Deterministic solution (b) Stochastic solution

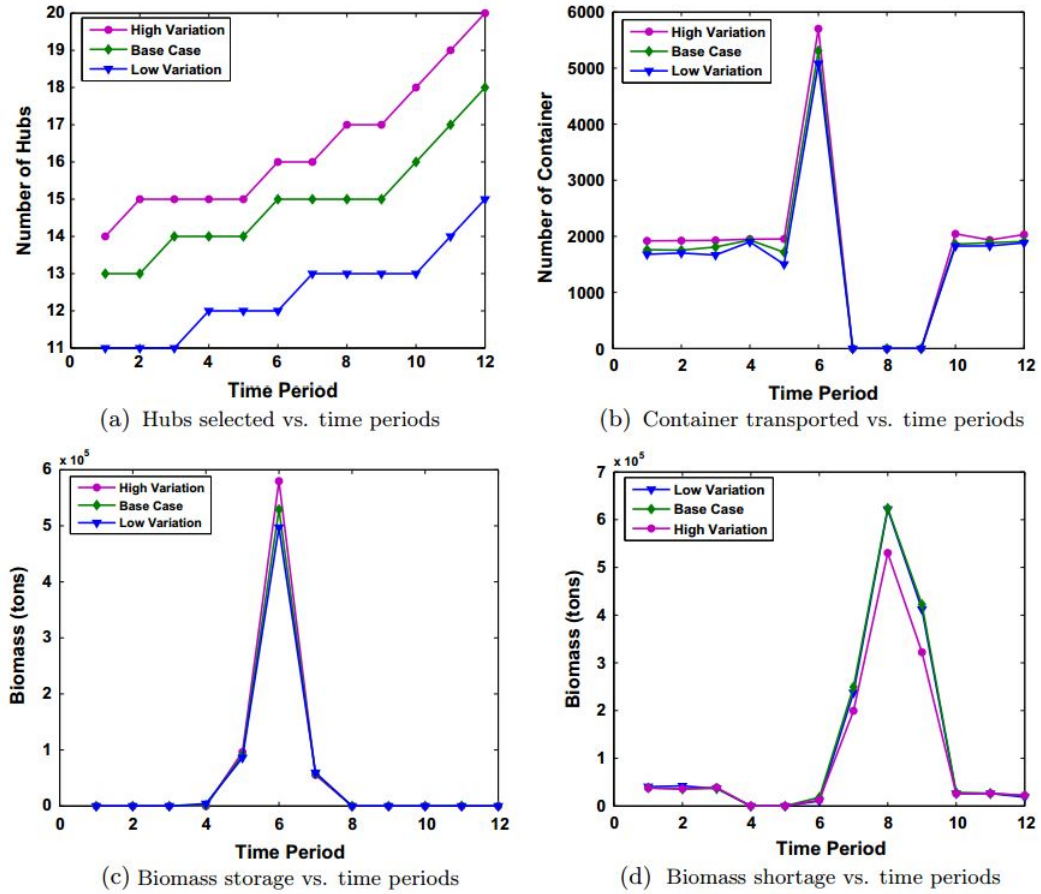
Figure 3.6

Biomass supply chain network configuration with and without considering biomass supply uncertainty

now tend to store more biomass during the off production seasons (December to February) when neither corn stover nor woody biomass is available. This in turn decreases the overall shortage quantity of biomass supply in each coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$. Figure 3.8 shows the distribution of multi-modal facilities under low and high biomass supply variability levels. It is observed that compared to low biomass supply variability levels, an additional six multi-modal facilities are selected to transport biomass in coal plants in high biomass supply variability levels which further increases the unit cost of biomass to $42.84/ton from $40.97/ton. This supports that the biomass supply variability levels highly impacts the biomass supply chain network decisions.

3.3.3.3 Impact of mean biomass supply changes on system performance:

The third set of experiments highlight the impact of biomass supply mean changes in supply chain network configuration. The mean biomass supply may change due to



(a) Hubs selected vs. time periods

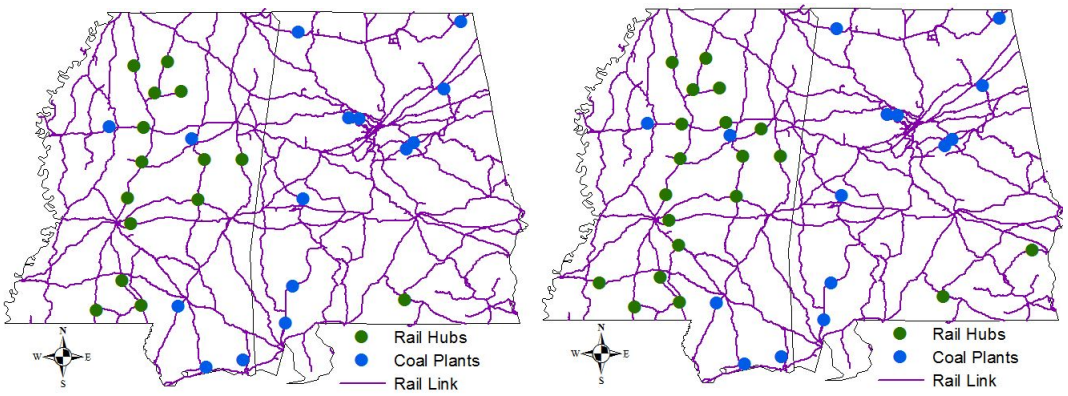
(b) Container transported vs. time periods

(c) Biomass storage vs. time periods

(d) Biomass shortage vs. time periods

Figure 3.7

Supply chain network decisions under different biomass supply variability levels



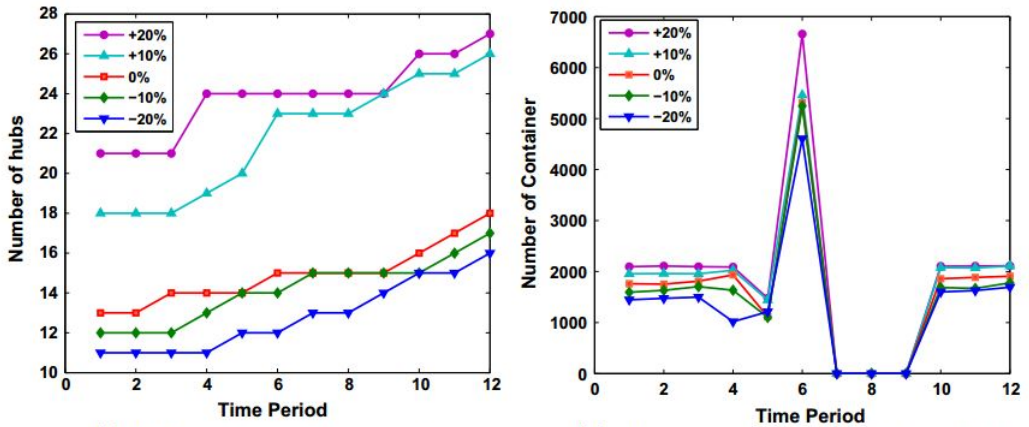
(a) Low supply variability

(b) High supply variability

Figure 3.8

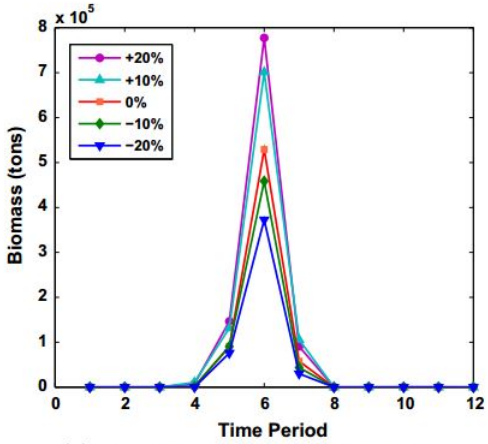
Impact of biomass supply variability on network configuration

unexpected natural catastrophe, new initiative from government (e.g., promoting farmers to cultivate more biomass) or any other related factors [73]. However, the biomass supply variability still exists from one time period to another despite the change in mean supply for biomass. Therefore, we now conduct some additional experiments to observe the change in mean biomass supply on supply chain network configuration. To run these experiments, we generate scenarios from the same variability level which are obtained from Section 3.2.1. Figure 3.9 presents the impact of biomass supply mean changes on supply chain network configuration. We observe that as the value of mean biomass supply $\overline{s}_{bit}, \forall b \in \mathcal{B}, i \in \mathcal{I}, t \in \mathcal{T}$ increases by 20%, model **[CSC]** selects an additional 49.8% multi-modal facilities to cope with high biomass mean supply changes (shown in Figure 3.10b). Furthermore, it is observed that the number of containers transported between the multi-modal facilities and the biomass storage quantities in coal plants is significantly impacted by biomass supply changes. This in turn drops the shortage quantity (**U**) of biomass supply at each coal plant $k \in \mathcal{K}$ by 25.9%. On the other hand, when we assume that the mean biomass supply ($\overline{s}_{bit}$) is decreased by 20%, the number of multi-modal facilities ($|\mathbf{Y}|$) as well as the number of containers ($|\mathbf{Z}|$) transported between link $(j, k) \in \mathcal{A}_2$ is decreased by 13.9% and 16.3%, respectively. A layout of the network configuration under this scenario is shown in Figure 3.10a. The low mean biomass supply scenario (e.g., -20% biomass supply change) increases the unit delivery cost of biomass to $43.54/ton from $40.97/ton which could have dropped down to $37.89/ton if there exists a high mean biomass supply scenario (e.g., 20% biomass supply change). In summary, it is observed that the mean biomass supply changes will highly impact the biomass supply chain network decisions.

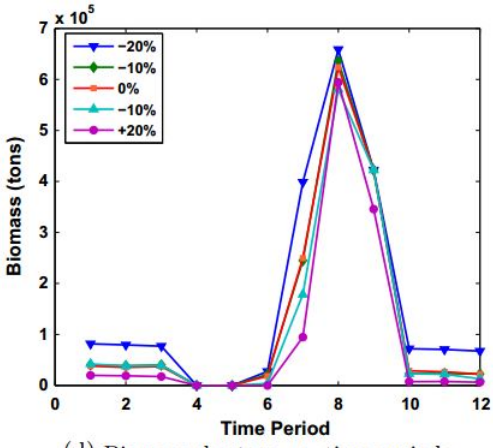


(a) Hubs selected vs. time periods

(b) Container transported vs. time periods



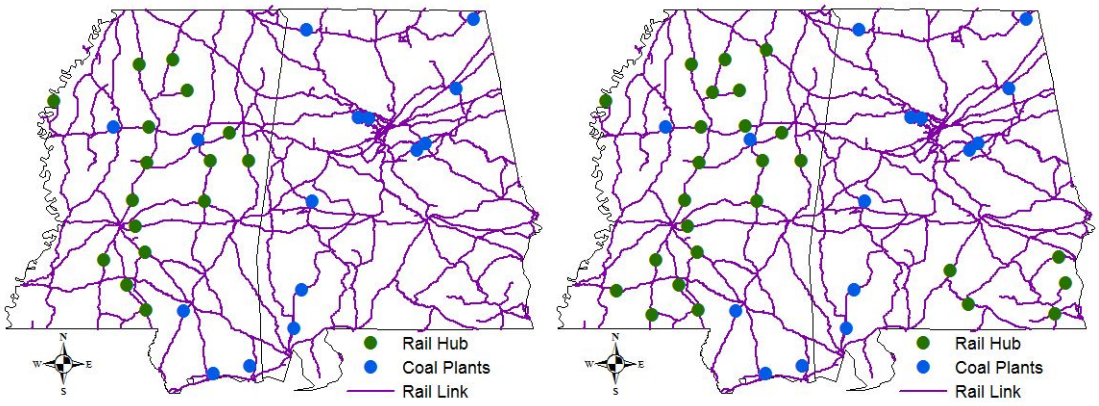
(c) Biomass storage vs. time periods



(d) Biomass shortage vs. time periods

Figure 3.9

Biomass supply chain network configuration under different supply chain mean levels



(a) Low biomass supply scenario (b) High biomass supply scenario

Figure 3.10

Impact of mean biomass supply changes on network configuration

3.3.3.4 Impact of Multi-modal Transportation on System Performance:

The fourth set of experiment highlights the impact of multi-modal facilities in biomass supply chain network configuration. We run a set of experiments considering truck to be the only mode to transport feedstock from biomass supplier sites to the coal plants and compare it with the case when both modes of transportation (i.e., rail and truck) are available. We compare the case under different biomass supply variation levels. To achieve this goal, we create two different scenarios: high and low biomass variations (as discussed previously in Section 4.3.2) and calculate the unit cost under the assumption that no multi-modal facilities are used; hence, everything will be shipped via highways.

Figure 3.11 shows the impact of not considering multi-modal facilities on biomass supply chain network. We observe that as the multi-modal facilities are not considered

the level of shortage of feedstock increases by 90.15% to 104.13% during low and high biomass variation levels, respectively. The shortage in the feedstock was fulfilled by purchasing the biomass from outside market at higher price (π_{kt}) of \$54/ton.

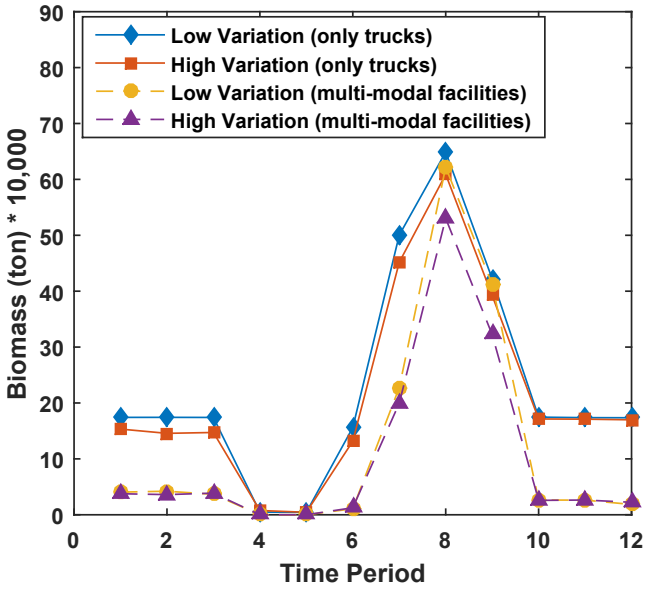


Figure 3.11

Biomass shortage vs. time periods under different supply variability level

Figure 3.12 shows changes of unit cost of biomass under both deterministic and stochastic settings and under the case when only trucks are used as the only modes of transportation to transport biomass from the feedstock supplier sites to the coal plants. It is observed that the unit cost of biomass increases to \$51.24/ton from \$45.11/ton under deterministic settings and \$49.74/ton from \$40.97/ton (low biomass variation) and \$50.39/ton from \$42.84/ton (high biomass variation) under stochastic settings and when only trucks are used. Results indicate that introducing multi-modal facilities in the biomass supply chain

network hedges against supply variations compared to using standalone truck transportation.

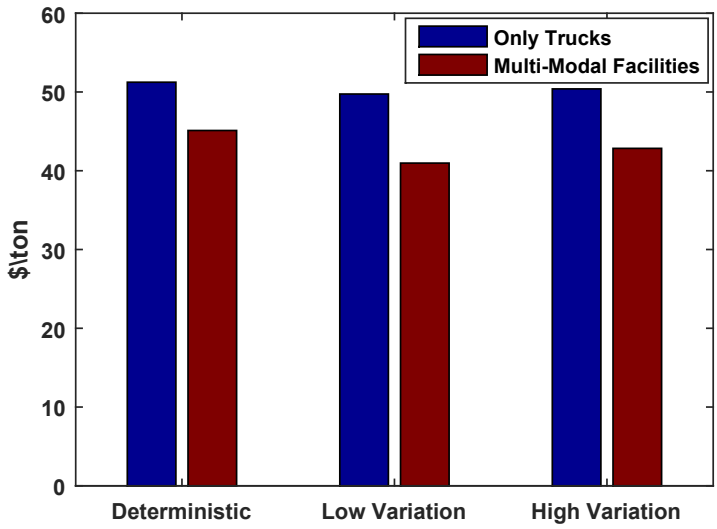


Figure 3.12

Unit cost per ton of biomass under different supply variability level

3.4 Conclusion

This paper studies the impact of feedstock supply uncertainty to the design and management of an inbound biomass co-firing supply chain network. A two-stage stochastic mixed integer linear programming model **[CSC]** is developed to determine the optimal use of multi-modal facilities, biomass storage and processing plans, and shipment routes for delivering biomass to coal plants under feedstock supply uncertainty. To represent a more realistic case, we generate a scenario tree based on the prediction errors obtained from historical and forecasted feedstock supply availability. The scenario tree problem is then

solved using a Sample average approximation algorithm that uses an enhanced Progressive hedging algorithm to solve the scenario subproblems. The enhanced Progressive hedging algorithm incorporates several algorithmic improvements such as the dynamic penalty parameter updating technique, the local and global heuristics techniques, and the Rolling horizon algorithm. Computational results indicate that the hybrid decomposition algorithm (**[SAA+PHA+HEU+RHA]**) is capable of producing high quality solutions in a reasonable amount of time to solve realistic large-size problem instances.

By using the states of Mississippi and Alabama as a testing ground for our study, we conducted thorough computational experiments to test our model and to draw managerial insights. Our computational experiments reveal some insightful results about the impact of feedstock supply uncertainty on a biomass co-fired supply chain network. Based on the results, it is observed that high feedstock supply variability increases the unit delivery cost of biomass to $42.84/ton from $40.97/ton. Moreover, when the mean feedstock supply increases (e.g., 20%) it will drop the unit delivery cost of biomass to $37.89/ton from $40.97/ton. The sensitivity analysis further reveals the justification of using stochastic programming model solutions over deterministic equivalent model solutions in solving model **[CSC]**.

In summary, the contributions of the paper to the literature are manifold. First, we present a two-stage stochastic programming model formulation that captures the details of inbound transportation network design decisions for a biomass co-firing supply chain under feedstock supply uncertainty. Second, we generate realistic scenarios by using the prediction errors obtained from historical and forecasted feedstock supply availability. The

scenario tree problem is then solved with high quality and in a time efficient manner by using a hybrid decomposition algorithm that connects Sample average approximation algorithm with an enhanced Progressive hedging algorithm. Finally, a real-world case study of the model is presented that reveals the impact of feedstock supply uncertainty on a biomass co-firing supply chain network designing problem. The findings can be used by decision makers to design a reliable inbound logistics networks for biomass co-firing supply chain.

This work can be extended in several directions. This study assumes that the multi-modal facilities are reliable and will never fail. However, in reality multi-modal facilities can be disrupted due to natural (e.g., 2005 Hurricane Katrina, 2008 China and 2009 Haiti Earthquakes) or human-made disasters (e.g., 2010 Gulf of Mexico Oil Spill). Furthermore, our work can be extended to consider congestions caused by feedstock seasonality in multi-modal facilities. These issues will be addressed in future studies.

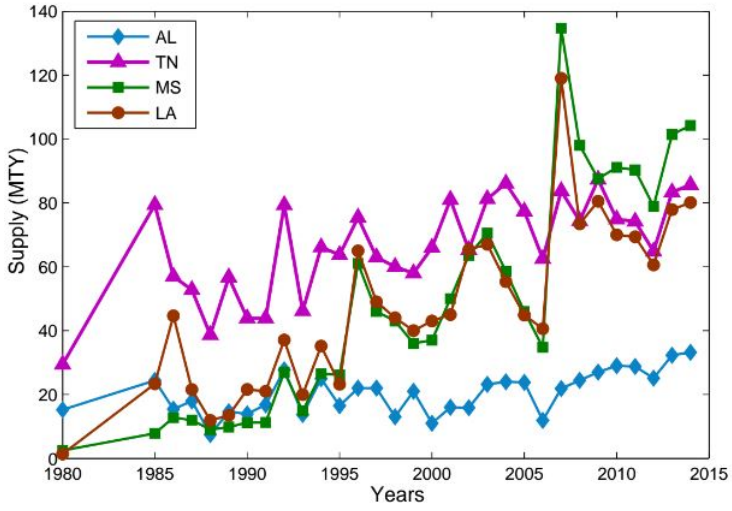


Figure 3.13

1980-2014 Cornstover supply fluctuation [111]

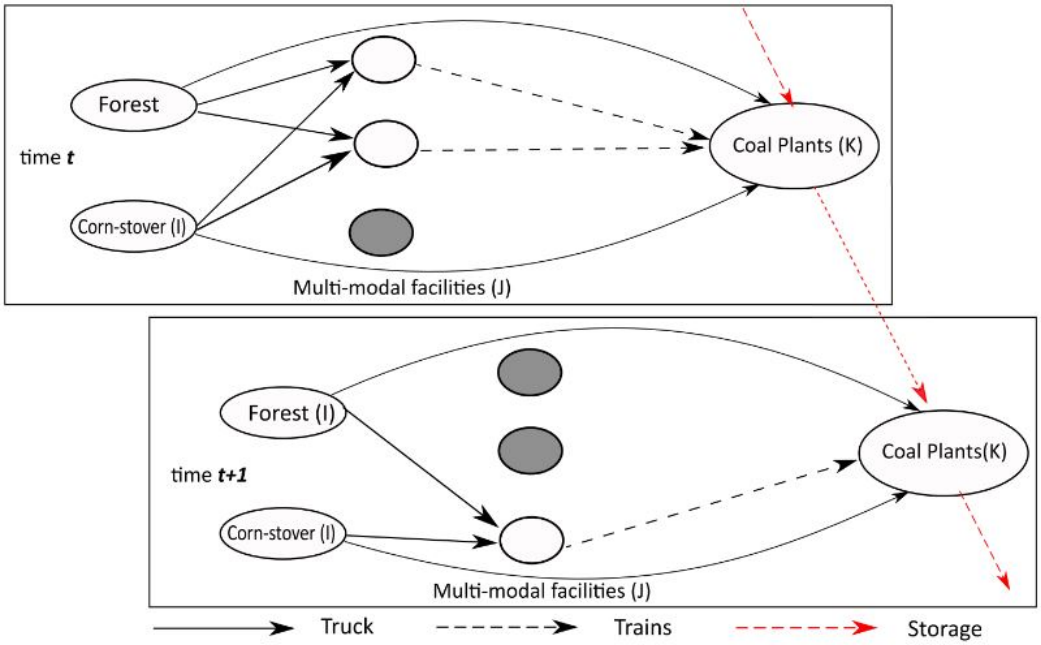


Figure 3.14

The biomass co-firing supply chain network

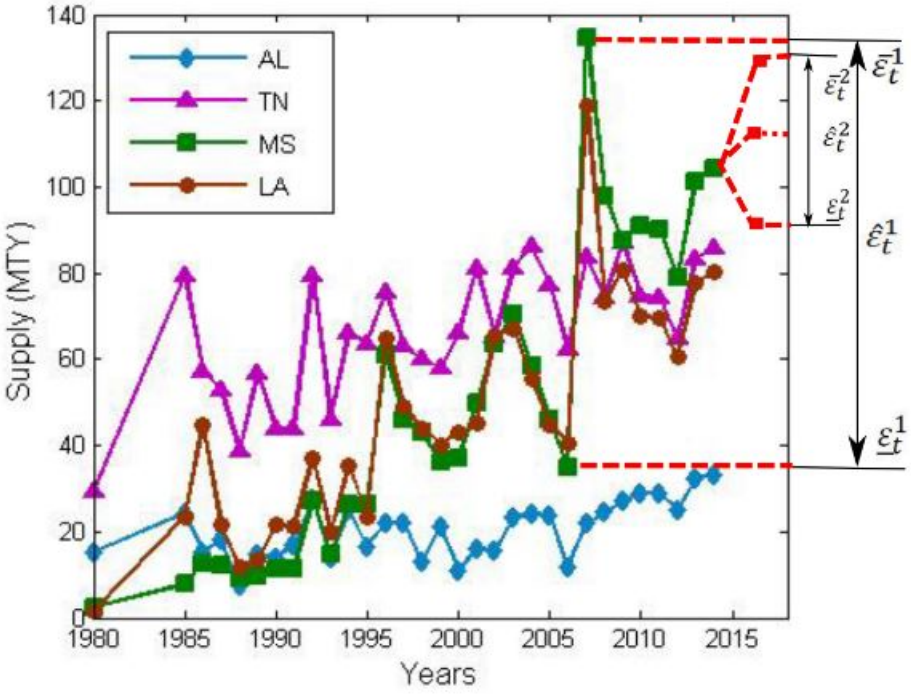


Figure 3.15

Biomass supply prediction for year 2015

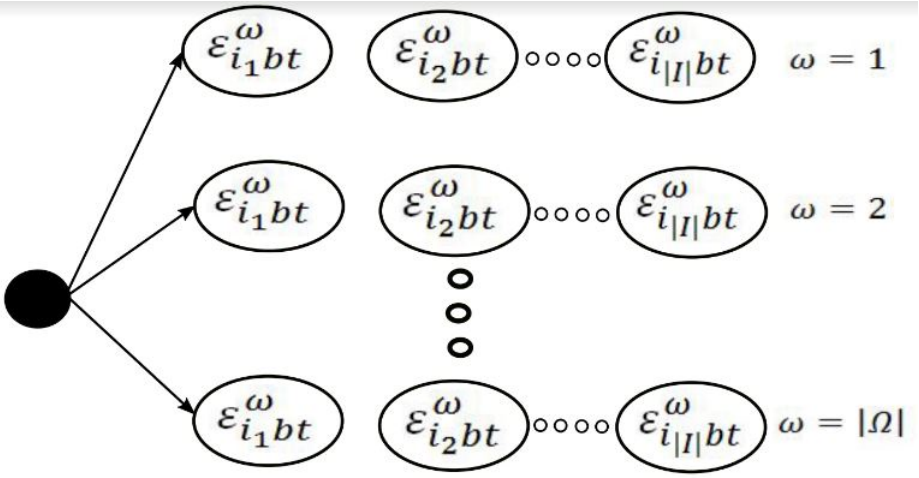


Figure 3.16

Scenario tree for the prediction error term

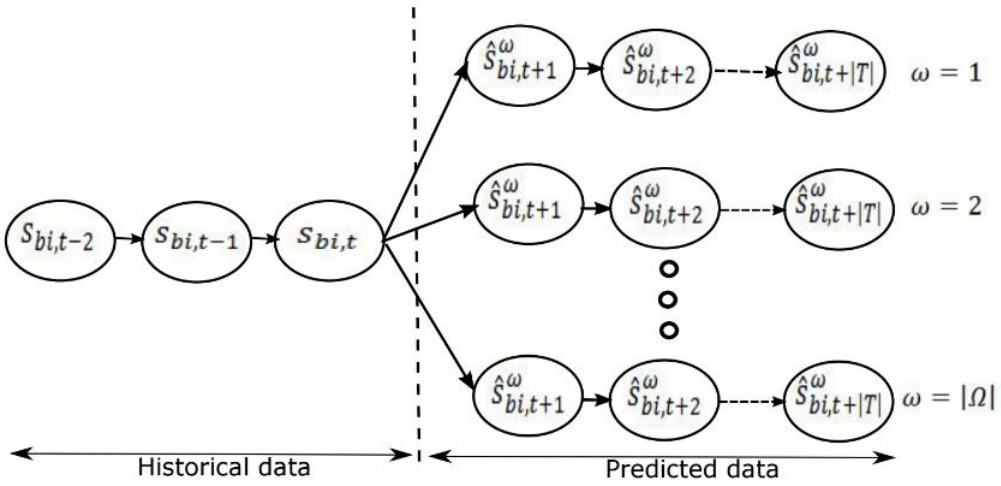


Figure 3.17

Biomass supply forecasting and scenario generation scheme

Table 3.3

Performance of the enhancement techniques used in PHA

N	$\|\mathcal{T}\|$	**[PHA]**			**[PHA+HEU]**			**[PHA+HEU+RHA]**		
		Gap (%)	CPU (sec)	Iter	Gap (%)	CPU (sec)	Iter	Gap (%)	CPU (sec)	Iter
Normal distribution										
20	4	9.15	10,800	35	0.91	2,375	7	0.92	**1,584**	3
	8	6.23	10,800	34	0.99	6,662	9	0.83	**2,894**	2
	12	66.38	10,800	11	0.95	7,562	6	0.89	**5,741**	4
30	4	9.86	10,800	24	0.95	3,467	6	0.91	**3,301**	4
	8	2.13	10,800	30	0.82	6,841	6	0.83	**4,122**	4
	12	65.41	10,800	8	0.35	9,428	5	0.74	**7,628**	3
40	4	4.24	10,800	17	0.97	**4,141**	6	0.89	4,645	5
	8	34.29	10,800	8	5.22	10,800	8	0.68	**6,287**	2
	12	47.32	10,800	5	3.23	10,800	5	0.55	**9,875**	2
Average		22.22	10,800	19.1	1.60	6,897	6.4	0.80	**5119.7**	3.2
Uniform distribution										
20	4	48.55	10,800	33	0.90	3,444	10	0.74	**1,384**	2
	8	33.30	10,800	16	0.76	6,723	9	0.86	**2,941**	2
	12	38.85	10,800	10	0.51	10,620	8	0.91	**6,842**	4
30	4	48.67	10,800	23	0.74	4,141	8	0.79	**3,180**	5
	8	24.90	10,800	11	0.92	10,262	9	0.81	**4,522**	2
	12	65.47	10,800	7	1.31	10,800	6	0.57	**8,854**	3
40	4	49.28	10,800	17	0.80	5,761	8	0.97	**4,385**	5
	8	27.90	10,800	8	4.82	10,800	5	0.86	**6,181**	2
	12	65.35	10,800	6	10.82	10,800	4	**1.03**	10,800	2
Average		44.70	10,800	14.6	2.40	8,150	7.4	0.84	**5,454.3**	3.0

Table 3.4

Comparison of different solution approaches

N	M	[SAA]			[SAA+PHA]			[SAA+PHA+HEU]			[SAA+PHA+HEU+RHA]		
		Gap (%)	CPU (sec)	Iter	Gap (%)	CPU (sec)	Iter	Gap (%)	CPU (sec)	Iter	Gap (%)	CPU (sec)	Iter
Normal distribution													
20	5	2.32	10,800	1	2.45	10,800	2	0.32	4,625	2	0.88	**4,522**	2
	10	*out of time*[a]			3.83	10,800	2	0.93	7,526	2	0.96	**6,904**	2
30	5	*out of memory*[b]			2.58	10,800	2	0.89	7,811	1	0.91	**5,721**	1
	10	*out of memory*			25.31	10,800	1	3.48	10,800	1	0.78	**9,562**	1
40	5	*out of memory*			*out of memory*			0.95	9,964	2	0.93	**7,941**	2
	10	*out of memory*			*out of memory*			3.51	10,800	1	0.94	**10,567**	1
Average		2.32	10800	1.0	8.54	10800	1.8	1.68	8588	1.5	0.90	**7536**	1.5
Uniform distribution													
20	5	0.65	8,482	1	0.64	4,689	1	0.94	3,871	2	0.96	**3,481**	2
	10	*out of time*			0.35	8,647	2	0.98	**5,126**	1	1.12	4,529	1
30	5	*out of memory*			0.86	9,240	1	0.90	7,861	2	0.94	**6,314**	2
	10	*out of memory*			*out of memory*			4.26	10,800	1	**1.38**	10,800	1
40	5	*out of memory*			*out of memory*			2.52	10,800	1	0.98	**8,115**	1
	10	*out of memory*			*out of memory*			6.81	10,800	1	**1.58**	10,800	1
Average		0.65[c]	8482	1.0	0.62[c]	7525	1.3	2.74	8210	1.3	1.16	**7340**	1.3

[a]unable to find any feasible solution within the time limit

[b]CPLEX runs out of memory

[c]Instances with (a) and (b) did not contribute to average calculations

CHAPTER 4

MANAGING CONGESTION IN A MULTI-MODAL TRANSPORTATION NETWORK UNDER BIOMASS SUPPLY UNCERTAINTY

4.1 Introduction

Biomass co-firing is a low-cost, sustainable option for efficiently and cleanly converting biomass to electricity. Biomass co-firing is predicted to have a sharp increase at coal-fired power plants. The Energy Information Administration (EIA) projected a sharp increase in biomass co-firing till 2040 [113] (shown in Figure 4.1). In fact, 40 out of 560 coal-fired plants in U.S. are already co-firing biomass with coal and this number is expected to increase in coming years. Such an increase in biomass co-firing mandates designing a lean transportation network that efficiently connects origin fields with destination coal plants with proper transportation modes. To achieve this goal, this study explores the key logistical challenges associated with biomass transportation and develops an optimization framework that helps decision makers better manage their supply chain network.

Switching from fossil fuels to renewable sources of energy is one of the best approaches to control Green House Gas (GHG) emissions in the energy supply sector. Direct co-firing is one of the most interesting and effective means of reducing GHG emissions from the coal-fired power plants. The supplies for co-firing potentially include woody biomass, short-rotation crops, agricultural residues and wastes, wood wastes, municipal

solid wastes, saw dust, grass, waste from food processing, animal wastes and a host of other materials [27]. There are some challenges associated with biomass co-firing that comes along with its benefits: (1) ash fusion temperature for biomass are lower than that for the coal ash fusion which will increase the rate and extend of boiler slagging; (2) higher moisture level as much as 50% moisture reduces the efficiency of boiler; (3) biomass are bulky and difficult to transport; (4) supply is uncertain and highly seasonal; and (5) biomass are widely spread geographically. Long haul transportation modes such as rail and barge can be utilized to transport biomass from supplier sites to coal plants and provide a means of solution against challenges (3) and (5) listed above. However, the uncertainty in supply and the seasonality will impact the operation of multi-modal facilities directly. For instance, the harvesting season for corn-stover follows the harvesting of corn, which starts from early September and ends in November. On the other hand, woody biomass is available all year round, except three months during the winter. Early September through late November is the peak biomass production season in U.S. when more biomass is expected to ship through multi-modal facilities. This seasonality not only congests the facilities for a given time period of the year, but also impacts overall biomass supply chain activities. Additionally, multi-modal facilities such as railways requires significant investment to alleviate the impact of congestion from the network. For example, Burlington Northern Santa Fe Corporation (BNSF) is planning to invest a record of $60 billion in 2015 to relieve network congestion caused by surging demand for grain and oil shipments [51]. A majority of this investment will go for expansion projects.

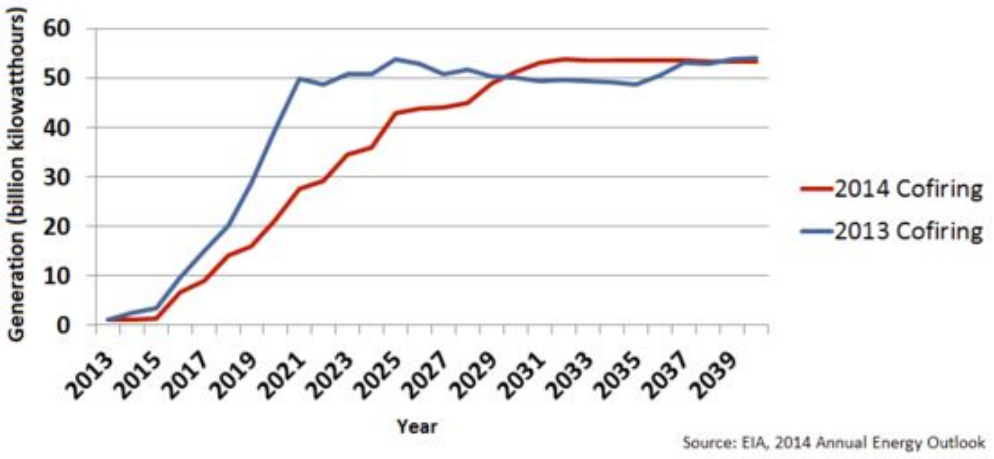
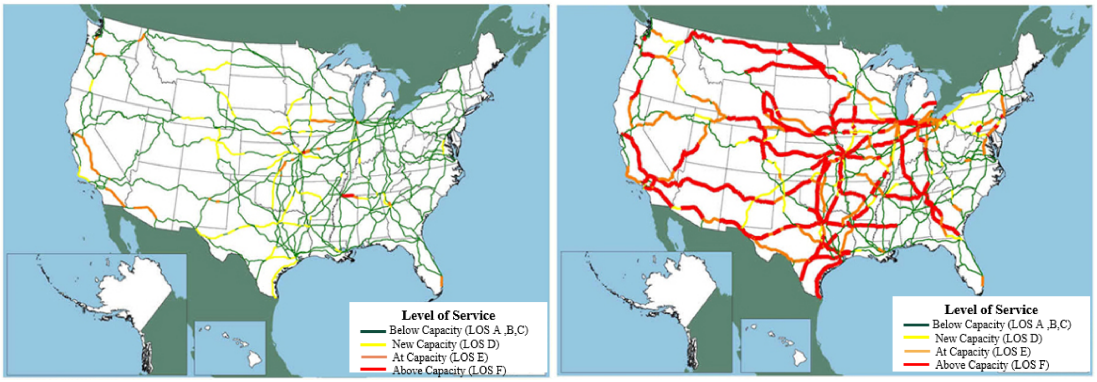


Figure 4.1

Biomass co-firing estimates of 2013 and 2014 [113]

U.S. railways are experiencing a tremendous increase in grain transportation for the last few years. The number has increased by 25% compared to the average of the last three years (2011-2013) grain transportation [113]. The spike in grain transportation not only causes the delay in transportation but also contributes other ripple effects such as increase road traffic and elevate levels of air pollution. Seasonal surges in freight demand and disruptions from incidents and maintenance activities add to congestion as volume reaches capacity on the reduced mainline railroad network. Figure 4.2 shows the comparison of train volume to capacity of railway where Figure 4.2a shows the condition at year 2007 and Figure 4.2b shows the condition at year 2035 if the demand keeps increasing based on a constant rate of 2007 while capacity remains unchanged. It is observed that the congestion of the mainline railroad network is forecasted to spread significantly by year 2035. According to Association of American Railroads (AAR), 30% of the network will experience unstable flows and service breakdown because of congestion in 2035 if no planning takes

place now [2]. This alarming fact encourages us to incorporate the effect of congestion in our proposed optimization model and plans to mitigate its effect in the context of a biomass supply chain network.



(a) Current train volume to capacity (b) Train volume in 2035 to capacity

Figure 4.2

Train Volume Compared to Railway Capacity [2]

The study of the biomass supply chain optimization has increased rapidly in recent years. Studies conducted by Zamboni et al. [132], Eksioglu et al. [29, 30], Huang et al. [49], An et al. [1], Bai et al. [5], Xie and Ouyang [129], Zhang et al. [3], Roni et al. [100],[93] and Memisoglu and Uster [80] analyze plant location and transportation issues in biofuel supply chain networks under a deterministic setting. These studies are extended by Cundiff et al. [23], Kim et al. [59], Chen and Fan [17], Gebreslassie et al. [38], Awudu and Zhang [3], and Marufuzzaman et al. [75] to represent a more realistic case which captures system uncertainties (e.g., biomass supply, demand, technology, pricing) using optimization models. A brief overview of studies dedicated to uncertainty and sustain-

ability in a biofuel supply chain network can be found from a recent study by Awudu and Zhang [54]. The other stream of research in biofuel supply chain community considers network sustainability (e.g., Zamboni et al. [131], Giarola et al. [41], You et al. [130]) and facility reliability (e.g., Li et al. [66], Wang and Ouyang [122], Bai et al. [6], Marufuzzaman et al. [76], Marufuzzaman and Eksioglu [73], Poudel et al. [92]) which impact the biofuel supply chain network. Note that all the studies discussed above assume that the facilities will never get congested despite the fact that feedstock seasonality coupled with system uncertainty present in the biofuel supply chain network.

In the context of general network design, researchers have become increasingly interested in the effect of facility congestions on network performance. Grove and O'Kelly [43] are among the first researchers who investigate the relationship between hub-and-spoke networks and congestion by simulating the daily operations of a single assignment hub-and-spoke network. Kara and Tansel [56] develop a function to minimize the latest travel time and excessive delays at the hubs. Marianov and Serra [72] model the hub-and-spoke network as an M/D/c queuing network and solve the optimization problem using a Tabu search heuristic. Elhedhli and Hu [32] introduce a non-linear cost term in the objective function of an uncapacitated hub location design problem. In Elhedhli and Wu [33], the authors further extend the previous study to a capacitated hub-and-spoke network design problem. In both papers, the authors linearize the non-linear cost function by approximating a set of tangent hyper-planes and then use a Lagrangian relaxation algorithm to solve the problem in a reasonable amount of time. Most recently, Camargo et al. [12] model the congestion as a convex cost function and develop a mixed integer nonlinear program-

ming model for a multiple allocation hub-and-spoke network design problem. The authors successfully solve as many as 81 nodes by using a generalized Benders decomposition algorithm. Few studies such as Miranda et al. [82] and Vidyarthi and Jayaswal [118] focus on the impact of facility congestion under demand uncertainty. A brief overview of the hub location problems and solution methodologies can be found in a recent study by SteadieSeifi et al. [109].

Although a few researchers have already highlighted the importance of incorporating the congestion factor in a single facility or in a hub-and-spoke network structure, little work has been conducted till now to identify the impact of facility congestion from a supply chain point of view. There is a specific need to address multi-modal congestion issues in a biomass supply chain network since this industry is highly impacted by feedstock seasonality along with different sources of uncertainties (e.g., biomass supply, demand, technology, government policies). Bai et al. [4] first introduce the traffic congestion in the traditional facility location model to decide the optimal locations of refineries and the flow of biomass and ethanol in the transportation network. The authors propose a Lagrangian relaxation algorithm that is nested under a branch-and-bound framework to find a high-quality feasible solution in a reasonable amount of time. Finally, the authors conduct a number of sensitivity analyses to show the effects of highway congestion on biorefinery location design and supply chain costs. Most recently, Hajibabai and Ouyang [45] propose an integrated mathematical model that aims to minimize the total cost of facility construction, roadway capacity expansion, including highway links and railway segments, and biomass/biofuel transportation delay due to congestion. In order to solve the model in a reasonable amount

of time, the authors develop a Genetic Algorithm that integrates Lagrangian relaxation and convex combination algorithms. A real world case study is developed to show the effects of roadway expansion and traveler time value on biofuel supply chain design decisions.

4.1.1 Research Scope and Contributions

The key feature distinguishing this study from most existing work on biomass supply chain is the the integration of uncertainty of biomass supply with the impact of facility congestion and feedstock seasonality on a coal-fired power plant supply chain system. The aim is to minimize the overall system costs which include the costs of biomass transportation, processing, storage, delay, and investment to use multi-modal facilities. To accomplish this goal, we have developed a two-stage stochastic mixed-integer non-linear programming model that provides location of multi-modal facilities, number of containers transported between the multi-modal facilities, amount of biomass processed, stored, and transported from feedstock supply sites to coal plants under biomass supply uncertainty while preventing congestion to occur. We develop a hybrid decomposition algorithm that combines Constraint Generation algorithm with a Sample Average Approximation algorithm and an enhanced Progressive Hedging algorithm to solve the optimization model. We use few enhancement techniques to accelerate the performance of the Progressive Hedging algorithm, such as local and global adjustment strategies, dynamic penalty parameter updating techniques, and Rolling Horizon heuristics. Through multiple experiments we have shown that the hybrid decomposition algorithm is capable of producing high quality solution to solve the large-size problem instances of our model in a reasonable amount of time.

In addition to proposing the general model, another important contribution of this paper is applying this model to a real-world case study. We use the states of Mississippi and Alabama as a testing ground for our case study. The outcome of this study provides a number of managerial insights such as the deployment of multi-modal facilities, biomass stored, processed, and transportation decisions under different biomass supply variability levels and congestion prices, which can effectively aid decision makers to design a reliable and realistic biomass co-firing supply chain network. Finally, we show how the mean biomass supply changes impact the supply chain network configuration and justified the use of considering stochastic programming modeling approach over deterministic approach considering biomass seasonality and multi-modal congestion factor into account.

The remainder of this paper is organized as follows. Section 4.2 presents the model formulation. Section 4.3 introduces various solution approaches such as Constraint Generation (CG), Sample Average Approximation (SAA), Progressive Hedging Algorithm (PHA), and Rolling Horizon (RH) heuristics to efficiently solve the proposed optimization model. Section 4.4 conducts numerical experiments to verify the algorithm performance and to draw managerial insights. Section 4.5 concludes this paper and provides future research directions.

4.2 Problem Description and Model Formulation

This section presents a two-stage stochastic programming model formulation for the design and management of a biomass co-firing supply chain network while taking the stochastic nature of biomass supply along with the impact of congestion in the multi-modal

facilities into account. In the two-stage stochastic programming model formulation, the first-stage decides the location, capacity, and timing to use multi-modal facilities among candidate locations prior to the realization of any random events. After the first-stage decisions are made, the random events are realized and the second-stage decisions such as the cost of procuring, storing, transporting biomass from feedstock supply sites to coal plants, and managing congestions in multi-modal facilities are made. The objective is to minimize the first-stage and expected second-stage costs across all possible feedstock supply scenarios for the biomass co-firing supply chain network. Figure 4.3 presents the structure of the supply chain network consisting of two biomass suppliers, three multi-modal facilities, and two coal plants.

4.2.1 Nonlinear Model Formulation

Consider a logistics network $\mathbb{G}(\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of arcs. Set $\mathcal{N}$ consists of the set of harvesting sites $\mathcal{I}$, the set of candidate multi-modal facilities $\mathcal{J}$, and the set of coal plants $\mathcal{K}$, i.e., $\mathcal{N} = \mathcal{I} \cup \mathcal{J} \cup \mathcal{K}$. The set of arcs $\mathcal{A}$ consists of three disjoint subsets, i.e., $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$, where $\mathcal{A}_1$ represents the set of arcs joining harvesting sites $\mathcal{I}$ with multi-modal facilities $\mathcal{J}$; $\mathcal{A}_2$ represents the set of arcs between multi-modal facilities $\mathcal{J}$ and coal plants $\mathcal{K}$; and finally, $\mathcal{A}_3$ represents the set of arcs that directly connects harvesting sites $\mathcal{I}$ to coal plants $\mathcal{K}$. Let $\mathcal{B}$ denotes the set of feedstock types to be distributed in the logistics network in different time periods of a year. We further let $\mathcal{T}$ be the set of time periods and $\mathcal{L}$ be the set of sizes for the multi-modal

facilities to be considered in this problem. Finally, we denote Ω as the sample space of the random event where $\omega \in \Omega$ defines a particular realization.

Let $s_{bit\omega}$ denotes the amount of biomass of type $b \in \mathcal{B}$ available at site $i \in \mathcal{I}$ in time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$. Each coal plant $k \in \mathcal{K}$ demands d_{kt} tons of biomass in time period $t \in \mathcal{T}$. We assume that the unmet demand for biomass can be substituted. The per unit cost of biomass substitution at the coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$ is denoted by π_{kt}. This will be the threshold cost that customers want to pay for the biomass. If the unit delivery cost of biomass exceeds this threshold cost, it will be beneficial to use the substitution rather than getting biomass from the internal supply chain network.

For nodes $j \in \mathcal{J}$, ψ_{ljt} denotes the fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at time period $t \in \mathcal{T}$. The multi-modal facilities already existing in the tested regions are considered as candidate locations for this study and the fixed cost represents the additional cost required to use the facilities (e.g., building additional tracks, purchasing lifts, track switch). Each arc $(i, j) \in \mathcal{A}_1$ carries biomass of type $b \in \mathcal{B}$ from a harvesting site i to a multi-modal facility j and are generally located close to each other (e.g., 10-20 miles). Therefore, trucks are preferred on $(i, j) \in \mathcal{A}_1$ and its unit transportation cost is represented by c_{bijt}. Arcs $(j, k) \in \mathcal{A}_2$ connect multi-modal facilities with coal plants and usually carries large-volume of biomass. Thus, long-haul transportation mode such as rail can be used as a major carrier between them. We represent c_{bjkt} as the unit transportation cost along arc $(j, k) \in \mathcal{A}_2$. Therefore, a unit flow along an origin-destination route $\{(i, j), (j, k)\}$ for biomass of type $b \in \mathcal{B}$ at time period $t \in \mathcal{T}$ costs $c_{bijkt} = c_{bijt} + c_{bjkt}$. Biomass is usually transported in cargo containers between the multi-modal facilities. We

impose a fixed cost ξ_{jkt} as the cost for loading and unloading cargo containers in the multi-modal facilities. Furthermore, this study allows biomass to be shipped directly by trucks along arcs $(i, k) \in \mathcal{A}_3$, which incurs a unit transportation cost c_{bikt}. These arcs are preferred when the harvesting sites $i \in \mathcal{I}$ are located close to the coal plants $k \in \mathcal{K}$ and thus direct shipments of biomass using trucks are cheaper compared to multi-modal facilities. Finally, biomass are required to be purchased from the farmers and collected and stored before hauling it on truck. Therefore, we denote δ_{bit} and γ_{bit} as the unit purchasing and storage cost for biomass of type $b \in \mathcal{B}$ at location $i \in \mathcal{I}$ and under time period $t \in \mathcal{T}$. Following notations summarize the sets and input parameters for our two-stage stochastic programming model formulation:

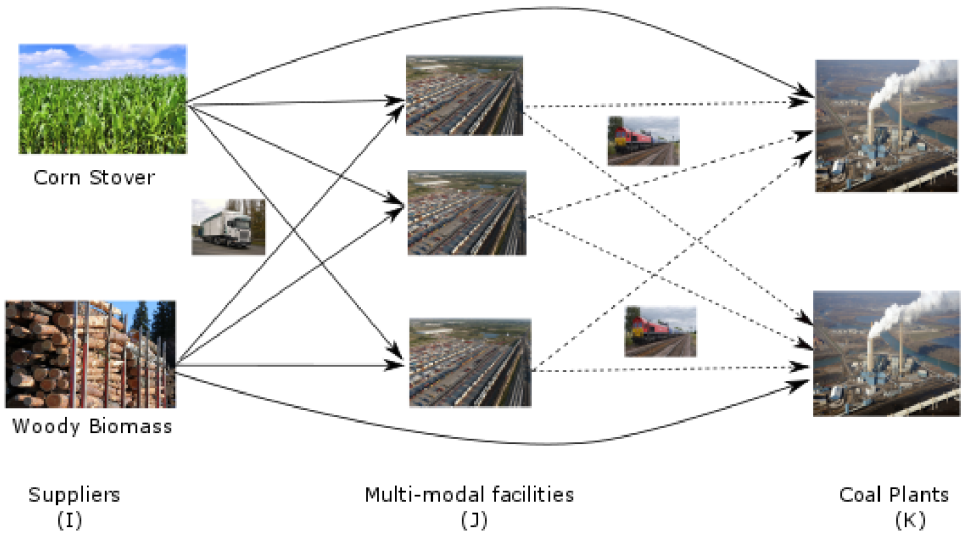


Figure 4.3

Supply chain network for biomass co-firing with coal

Sets:

- $\mathcal{I}$: set of harvesting sites (farms)
- $\mathcal{J}$: set of multi-modal facilities
- $\mathcal{K}$: set of coal plants
- $\mathcal{L}$: set of facility capacities
- $\mathcal{T}$: set of time periods
- $\mathcal{B}$: set of biomass types (b_1 for corn-stover and b_2 for forest residues)
- Ω: set of scenarios

Parameters:

- ψ_{ljt}: fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$
- ξ_{jkt}: fixed cost of a cargo container for transporting biomass along arc $(j,k) \in \mathcal{A}_2$ in time period $t \in \mathcal{T}$
- c_{bikt}: unit transportation cost for biomass of type $b \in \mathcal{B}$ along arc $(i,k) \in \mathcal{A}_3$ in time period $t \in \mathcal{T}$
- c_{bijkt}: unit transportation cost for biomass of type $b \in \mathcal{B}$ along arc $(i,j,k) \in \mathcal{A}_1 \cup \mathcal{A}_2$ in time period $t \in \mathcal{T}$
- $s_{bit\omega}$: amount of biomass of type $b \in \mathcal{B}$ available at site $i \in \mathcal{I}$ in period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- d_{kt}: demand for biomass at coal plant $k \in \mathcal{K}$ in period $t \in \mathcal{T}$
- c^{cap}: capacity of cargo container
- h_k^{cap}: biomass holding capacity at coal plant $k \in \mathcal{K}$
- C_{lj}: biomass storage/handling capacity of a multi-modal facility of size $l \in \mathcal{L}$ at location $j \in \mathcal{J}$
- α_b: deterioration rate of biomass of type $b \in \mathcal{B}$
- p_{bkt}: unit biomass processing cost of type $b \in \mathcal{B}$ at coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$

- h_{bkt}: unit inventory holding cost for biomass of type $b \in \mathcal{B}$ at coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$
- δ_{bit}: unit biomass purchasing cost of type $b \in \mathcal{B}$ at location $i \in \mathcal{I}$ in time period $t \in \mathcal{T}$
- γ_{bit}: unit biomass storage cost of type $b \in \mathcal{B}$ at location $i \in \mathcal{I}$ in time period $t \in \mathcal{T}$
- π_{kt}: unit penalty cost of not satisfying demand at coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$
- ρ_{ω}: probability of scenario $\omega \in \Omega$

Decision Variables:

- Y_{ljt}: 1 if a multi-modal facility $j \in \mathcal{J}$ of capacity $l \in \mathcal{L}$ is used at time period $t \in \mathcal{T}$; 0 otherwise
- $Z_{jkt\omega}$: number of cargo containers transported between multi-modal facility $j \in \mathcal{J}$ to coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $X_{bikt\omega}$: amount of biomass of type $b \in \mathcal{B}$ transported from harvesting site $i \in \mathcal{I}$ to coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $X_{bijkt\omega}$: amount of biomass of type $b \in \mathcal{B}$ transported from harvesting site $i \in \mathcal{I}$ to coal plant $k \in \mathcal{K}$ through multi-modal facility $j \in \mathcal{J}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $P_{bkt\omega}$: amount of biomass of type $b \in \mathcal{B}$ processed in coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $H_{bkt\omega}$: amount of inventory of type $b \in \mathcal{B}$ stored in coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$
- $U_{kt\omega}$: shortage of biomass demand at coal plant $k \in \mathcal{K}$ in period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$

We now introduce the first and second-stage decision variables for our two-stage stochastic programming model formulation. The first-stage decision variables $\mathbf{Y} := \{Y_{ljt}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ determine the size, location, and time to use a multi-modal facilities, i.e.,

$$Y_{ljt} = \begin{cases} 1 & \text{if a multi-modal facility of size } l \text{ is used in location } j \text{ at time period } t \\ 0 & \text{otherwise;} \end{cases}$$

The second-stage decision variables $\mathbf{Z} := \{Z_{jkt\omega}\}_{j\in\mathcal{J},k\in\mathcal{K},t\in\mathcal{T},\omega\in\Omega}$ determines the number of containers transported between the multi-modal facilities $j \in \mathcal{J}$ to coal plants $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$; $\mathbf{X} := \{X_{bert\omega}\}_{b\in\mathcal{B},(e,r)\in\mathcal{A},t\in\mathcal{T},\omega\in\Omega}$ denote the flow of biomass of type $b \in \mathcal{B}$ along each link $(e, r) \in \mathcal{A}$ of the network at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$; $\mathbf{P} := \{P_{bkt\omega}\}_{b\in\mathcal{B},k\in\mathcal{K},t\in\mathcal{T},\omega\in\Omega}$ denote the amount of biomass of type $b \in \mathcal{B}$ processed in coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$; $\mathbf{H} := \{H_{bkt\omega}\}_{b\in\mathcal{B},k\in\mathcal{K},t\in\mathcal{T},\omega\in\Omega}$ denote the amount of biomass of type $b \in \mathcal{B}$ stored in coal plant $k \in \mathcal{K}$ at time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$; and $\mathbf{U} := \{U_{kt\omega}\}_{k\in\mathcal{K},t\in\mathcal{T},\omega\in\Omega}$ denote the amount of biomass shortage at coal plant $k \in \mathcal{K}$ in period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$.

Both corn-stover (b_1) and forest residues (b_2) are available during the peak biomass season which occurs during the months from September till November. Multi-modal facilities are expected to congest during this peak harvesting seasons of biomass which may lead to a dramatic increase in total supply chain cost. The impact of congestion becomes more severe when the total flow of feedstock $X_{bijkt\omega}$ approaches the capacity C_{lj} of a multi-modal facility $j \in \mathcal{J}$ [33]. Under steady-state conditions, the system-wide average waiting time

for the entire network can be represented as: $\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijkt\omega}}{\sum_{l\in\mathcal{L}}C_{ij}Y_{ljt}-\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijkt\omega}}\right)$. For a facility $j \in \mathcal{J}$, when the amount of feedstock $X_{bijkt\omega}$ increases, the ratio of this equation will also increase exponentially. Thus, the impact of congestion can be addressed realistically by the model. Now, let's consider c_0 be the congestion factor, then the system-wide congestion cost becomes: $\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}c_0\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijkt\omega}}{\sum_{l\in\mathcal{L}}C_{ij}Y_{ljt}-\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijkt\omega}}\right)$. Taking these factors into account, the following two-stage stochastic Mixed-Integer Nonlinear Programming (MINLP) model, referred to as **[CSM]**, can be formulated as follows:

$$\textbf{[CSM]}\ \underset{\mathbf{Y,Z,X,P,H,U}}{Minimize}\ \sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt}+\sum_{\omega\in\Omega}\rho_{\omega}\Bigg(\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt\omega}+\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{bikt\omega}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijkt\omega}+\sum_{k\in\mathcal{K}}p_{bkt}P_{bkt\omega}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bkt\omega}+\sum_{j\in\mathcal{J}}c_0\Bigg(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijkt\omega}}{\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt}-\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijkt\omega}}\Bigg)\Bigg)+\sum_{k\in\mathcal{K}}\pi_{kt}U_{kt\omega}\Bigg)\Bigg) \quad (4.1)$$

Subject to

$$\sum_{k\in\mathcal{K}} X_{biktω} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} X_{bijktω} \leq s_{bitω} \quad \forall b\in\mathcal{B}, i\in\mathcal{I}, t\in\mathcal{T}, ω\in\Omega \tag{4.2}$$

$$\sum_{i\in\mathcal{I}} X_{biktω} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} X_{bijktω} + (1-\alpha_b)H_{bk,t-1,ω} = H_{bktω} + P_{bktω} \; \forall b\in\mathcal{B}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{4.3}$$

$$\sum_{b\in\mathcal{B}} P_{bktω} + U_{ktω} = d_{kt} \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{4.4}$$

$$\sum_{b\in\mathcal{B}} H_{bktω} \leq h_k^{cap} \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{4.5}$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}} X_{bijktω} \leq c^{cap} Z_{jktω} \quad \forall (j,k)\in\mathcal{A}_2, t\in\mathcal{T}, ω\in\Omega \tag{4.6}$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijktω} \leq \sum_{l\in\mathcal{L}} C_{lj} Y_{ljt} \quad \forall j\in\mathcal{J}, t\in\mathcal{T}, ω\in\Omega \tag{4.7}$$

$$\sum_{l\in\mathcal{L}} Y_{ljt} \leq 1 \quad \forall j\in\mathcal{J}, t\in\mathcal{T} \tag{4.8}$$

$$Z_{jktω} \leq \sum_{l\in\mathcal{L}} \left\lceil \frac{C_{lj}}{c^{cap}} \right\rceil Y_{ljt} \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{4.9}$$

$$Y_{ljt} \in \mathbb{B} \quad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T} \tag{4.10}$$

$$Z_{jktω} \in Z^+ \quad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{4.11}$$

$$X_{bijktω}, X_{biktω}, H_{bktω}, P_{bktω}, U_{ktω} \geq 0 \; \forall b\in\mathcal{B}, i\in\mathcal{I}, j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \tag{4.12}$$

In **[CSM]**, the objective function minimizes the first-stage costs and the expected value of the random second-stage costs for the biomass co-firing supply chain network. The objective function consists of eight terms: the first term represents the costs of using multi-modal facilities in different time period of the year; the second term represents the

fixed costs of loading and unloading cargo containers between the multi-modal facilities; the third and forth term represent the expected transportation costs through highways and multi-modal facilities along with purchasing and storing costs of feedstock at the harvesting sites; the expected processing and storing costs at the coal plants are represented by the fifth and sixth terms of the objective function; the seventh term represents the expected costs due to congestion at the multi-modal facilities; and finally the last term of the objective function represents the expected penalty costs for biomass supply shortages.

Constraints (4.2) indicate that the amount of biomass of type $b \in \mathcal{B}$ supplied from supplier site $i \in \mathcal{I}$ in period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$ is limited by its availability. Constraints (4.3) are the flow balance constraints which ensure that at any time period $t \in \mathcal{T}$ and under scenario $\omega \in \Omega$ the biomass of type $b \in \mathcal{B}$ can be either produced or stored in a coal plant $k \in \mathcal{K}$. Constraints (4.4) indicate that the demand must be satisfied either through processed biomass obtained from the biomass supply chain network or through substitute products available in the market. Constraints (4.5) limit the amount of biomass that can be stored in a coal plant $k \in \mathcal{K}$ in time period $t \in \mathcal{T}$ to the maximum storage capacity of that plant h_k^{cap}; $\forall k \in \mathcal{K}$. Constraints (4.6) limit the amount of feedstock transported through arcs $(j,k) \in \mathcal{A}_2$ using multi-modal facilities at time $t \in \mathcal{T}$ under scenario $\omega \in \Omega$. Constraints (4.7) indicate that the total amount of biomass shipped through multi-modal facility $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$ under scenario $\omega \in \Omega$ is limited by the facility capacity C_{lj}; $\forall l \in \mathcal{L}, j \in \mathcal{J}$. Constraints (4.8) indicate that at most one multi-modal facility of capacity $l \in \mathcal{L}$ can be used at location $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$. Constraints (4.9) indicate that if no multi-modal facility is used at location $j \in \mathcal{J}$ then no

containers will flow between arcs $(j,k) \in \mathcal{A}_2$ at time period $t \in \mathcal{T}$ and under scenario $\omega \in \Omega$. Note that constraints (4.9) serve as valid-in-equalities for model **[CSM]** which helps the branch-and-bound process while solving the optimization problem. Finally, constraints (4.10) are the binary constraints, (4.11) are the integrity constraints, and (4.12) are the standard non-negativity constraints.

4.2.2 Model Linearization

Model **[CSM]** is nonlinear due to the presence of congestion function in the objective function of the model. We use the technique proposed by Elhedhli and Wu [33] to linearize the congestion term. Let's now introduce a new decision variable $\mathbf{M} := \{M_{jt\omega}\}_{j\in\mathcal{J},t\in\mathcal{T},\omega\in\Omega}$ as follows:

$$M_{jt\omega} = \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijkt\omega}}{\sum_{l\in\mathcal{L}} C_{lj}Y_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijkt\omega}} \tag{4.13}$$

Equation (4.13) can be further reduced as follow:

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijkt\omega} = \left(\frac{M_{jt\omega}}{1+M_{jt\omega}}\right)\sum_{l\in\mathcal{L}} C_{lj}Y_{ljt} = \sum_{l\in\mathcal{L}} C_{lj}\left(\frac{M_{jt\omega}}{1+M_{jt\omega}}\right)Y_{ljt}$$
$$\forall j \in \mathcal{J}, t \in \mathcal{T}, \omega \in \Omega \tag{4.14}$$

We now introduce another continuous variable $\mathbf{R} := \{R_{ljt\omega}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T},\omega\in\Omega}$ as follows:

$$R_{ljt\omega} = \left(\frac{M_{jt\omega}}{1+M_{jt\omega}}\right)Y_{ljt} \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}, \omega \in \Omega \tag{4.15}$$

Given that $Y_{ljt} = 1$, the above equation becomes:

$$\sum_{l \in \mathcal{L}} R_{ljt\omega} = \frac{M_{jt\omega}}{1 + M_{jt\omega}} \qquad \forall j \in \mathcal{J}, t \in \mathcal{T}, \omega \in \Omega \tag{4.16}$$

In the condition when $\{Y_{ijt}\}_{i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T}} = 0$ constraints (4.15) forces $\{R_{ljt\omega}\}_{l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}, \omega \in \Omega} = 0$. This is ensured by adding an additional constraint $0 \leq R_{ljt\omega} \leq Y_{ljt} \ \ \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}, \omega \in \Omega$ in the model formulation.

Lemma 2 *The function* $R_{ljt\omega}(M_{jt\omega}) = \frac{M_{jt\omega}}{1+M_{jt\omega}}$ *is concave in* $M_{jt\omega} \in (0, \infty)$.

Proof: While differentiating the function $R_{ljt\omega}$ w.r.t. $M_{jt\omega}$, we get the first derivative, $\frac{\delta}{\delta M_{jt\omega}}(R_{ljt\omega}) = 1/(1 + M_{jt\omega})^2 > 0$, and the second derivative, $\frac{\delta^2}{\delta M_{jt\omega}^2}(R_{ljt\omega}) = -2/(1 + M_{jt\omega})^3 < 0$. The first derivative is positive while the second derivative is negative; thus, it proves that the function $R_{ljt\omega}(M_{jt\omega})$ is concave in $M_{jt\omega}$. ∎

Lemma 1 implies that function $R_{ljt\omega}(M_{jt\omega})$ is concave and can be approximated by a set of tangent cutting planes as shown below [33]:

$$\frac{M_{jt\omega}}{1 + M_{jt\omega}} = Min_{h \in \mathcal{H}} \left[\frac{M_{jt\omega}}{\left(1 + M_{jt\omega}^h\right)^2} + \left(\frac{M_{jt\omega}^h}{1 + M_{jt\omega}} \right)^2 \right] \tag{4.17}$$

which is equivalent to:

$$\frac{M_{jt\omega}}{1 + M_{jt\omega}} \leq \frac{M_{jt\omega}}{\left(1 + M_{jt\omega}^h\right)^2} + \left(\frac{M_{jt\omega}^h}{1 + M_{jt\omega}} \right)^2 \qquad \forall j \in \mathcal{J}, t \in \mathcal{T}, h \in \mathcal{H}, \omega \in \Omega \tag{4.18}$$

where $\{M_{jt\omega}^h\}_{j \in \mathcal{J}, t \in \mathcal{T}, \omega \in \Omega, h \in \mathcal{H}}$ are the set of points used to approximate equation (4.18). The value of $\{Y_{ljt}\}_{l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}}$ is finite; therefore, the value that $\{M_{jt\omega}\}_{j \in \mathcal{J}, t \in \mathcal{T}, \omega \in \Omega}$ pro-

vides is also finite. This implies that the set $\mathcal{H}$ should be finite. We now derive equation (4.19) using (4.16) and (4.18) as follows:

$$\sum_{l\in\mathcal{L}} R_{ljt\omega} \leq \frac{M_{jt\omega}}{\left(1+M^h_{jt\omega}\right)^2} + \left(\frac{M^h_{jt\omega}}{1+M_{jt\omega}}\right)^2 \qquad \forall j\in\mathcal{J}, t\in\mathcal{T}, h\in\mathcal{H}, \omega\in\Omega \quad (4.19)$$

The linearized objective function **[LCSM]** for model **[CSM]** now becomes:

$$\textbf{[LCSM]} Minimize \sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt} + \sum_{\omega\in\Omega}\rho_\omega\Bigg(\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt\omega} +$$
$$\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\left(c_{bikt}+\delta_{bit}+\gamma_{bit}\right)X_{biktw} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\left(c_{bijkt}+\delta_{bit}+\gamma_{bit}\right)X_{bijktw} +$$
$$\sum_{k\in\mathcal{K}} p_{bkt}P_{bktw} + \sum_{k\in\mathcal{K}} h_{bkt}H_{bktw}\Bigg) + \sum_{j\in\mathcal{J}} c_0 M_{jtw} + \sum_{k\in\mathcal{K}}\pi_{kt}U_{ktw}\Bigg)\Bigg)$$

Subject to

$$\sum_{k\in\mathcal{K}} X_{biktω} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} X_{bijktω} \leq s_{bitω} \quad \forall b\in\mathcal{B}, i\in\mathcal{I}, t\in\mathcal{T}, ω\in\Omega \quad (4.20)$$

$$\sum_{i\in\mathcal{I}} X_{biktω} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} X_{bijktω} + (1-\alpha_b)H_{k,t-1,ω} = H_{bktω} + P_{bktω} \ \forall b\in\mathcal{B}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \quad (4.21)$$

$$\sum_{b\in\mathcal{B}} P_{bktω} + U_{ktω} = b_{kt} \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \quad (4.22)$$

$$\sum_{b\in\mathcal{B}} H_{bktω} \leq h_k^{cap} \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \quad (4.23)$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}} X_{bijktω} \leq v^{cap} Z_{jktω} \quad \forall (j,k)\in\mathcal{A}_2, t\in\mathcal{T}, ω\in\Omega \quad (4.24)$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijktω} \leq \sum_{l\in\mathcal{L}} C_{lj} R_{ljtω} \quad \forall j\in\mathcal{J}, t\in\mathcal{T}, ω\in\Omega \quad (4.25)$$

$$\sum_{l\in\mathcal{L}} R_{ljtω} \leq \frac{M_{jtω}}{\left(1+M_{jtω}^h\right)^2} + \left(\frac{M_{jtω}^h}{1+M_{jtω}}\right)^2 \quad \forall j\in\mathcal{J}, t\in\mathcal{T}, ω\in\Omega, h\in\mathcal{H} \quad (4.26)$$

$$R_{ljtω} \leq Y_{ljt} \quad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T}, ω\in\Omega \quad (4.27)$$

$$\sum_{l\in\mathcal{L}} Y_{ljt} \leq 1 \quad \forall j\in\mathcal{J}, t\in\mathcal{T} \quad (4.28)$$

$$Z_{jktω} \leq \sum_{l\in\mathcal{L}} \left\lceil \frac{C_{lj}}{C^{cap}} \right\rceil Y_{ljt} \ \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \quad (4.29)$$

$$Y_{ljt} \in \mathbb{B} \quad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T} \quad (4.30)$$

$$Z_{jktω} \in Z^+ \quad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \quad (4.31)$$

$$X_{bijktω}, X_{biktω}, H_{bktω}, P_{bktω}, U_{ktω}, R_{ljtω}, M_{jtω} \geq 0 \ \forall b\in\mathcal{B}, i\in\mathcal{I}, j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, ω\in\Omega \quad (4.32)$$

4.3 Solution Approach

This section presents the solution techniques used to solve model **[LCSM]**. Note that by setting $|\Omega| = 1$ and $|\mathcal{T}| = 1$ i.e., a single scenario and a single season, we can show that problem **[LCSM]** is a special case of a capacitated facility location problem which is known to be an $\mathcal{NP}$-hard problem [69]. Therefore, commercial solvers, such as CPLEX, cannot solve large-scale instances of this problem. In this section, we propose a hybrid decomposition algorithm that combines Constraint Generation algorithm, Sample Average Approximation algorithm, and enhanced Progressive Hedging algorithm to solve problem **[LCSM]**. The aim is to generate high quality solution for problem **[LCSM]** in a reasonable amount of time.

4.3.1 Constraint Generation Algorithm

In model **[LCSM]**, (4.26) generates large number of constraints. Therefore, it is really challenging to solve model **[LCSM]** while considering all the constraints at once. This motivates us to develop a Constraint Generation (CG) algorithm that can efficiently solve model **[LCSM]** despite generating large number of constraints through (4.26). The algorithm, extensively studied in [126] and [117], proceeds by solving a series of integer programs with a subset of the constraints obtained from (4.26) and added thereafter as needed. The process stops when the algorithm finds a solution for the sub-problem which does not violate any constraints within some accepted tolerance in the full problem **[CSM]**. Otherwise, a new set of points and thus a new set of constraints/cuts are generated which are added to **[LCSM]** in the next iteration. The algorithm is discussed in details below:

Let UB^q and LB^q denote an upper and lower bound of the original problem at iteration q. We further let ***v*[LCSM]** be the solution of the objective function value of **[LCSM]** and $(\mathbf{Y}^q, \mathbf{Z}^q, \mathbf{X}^q, \mathbf{P}^q, \mathbf{H}^q, \mathbf{U}^q)$ be its optimal solution. The following proposition provides the lower bound of the CG algorithm.

Proposition 1 *For any given subset of points* $\{M^h_{jt\omega}\}_{\mathcal{H}^q \subset \mathcal{H}}$, (4.33) *provides the lower bound of the optimal objective function value of* ***[CSM]****.*

$$LB = \textbf{\textit{v}[LCSM]}(\mathcal{H}^q) = \sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt} + \sum_{\omega\in\Omega}\rho_\omega\Bigg(\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt\omega} + \tag{4.33}$$

$$\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{bikt\omega} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijkt\omega} +$$

$$\sum_{k\in\mathcal{K}}p_{bkt}P_{bkt\omega} + \sum_{k\in\mathcal{K}}h_{bkt}H_{bkt\omega}\Bigg) + \sum_{j\in\mathcal{J}}c_0M_{jt\omega} + \sum_{k\in\mathcal{K}}\pi_{kt}U_{kt\omega}\Bigg)\Bigg)$$

Proof: **[LCSM]($\mathcal{H}^q$)** is the relaxed version of problem **[LCSM]**. Thus, the optimal objective function value ***v*[LCSM]**$(\mathcal{H}^q)$ obtained from equation (4.33) provides the lower bound to the optimal objective value of **[LCSM]**. Since problem **[LCSM]** is an approximation for problem **[CSM]**, solution ***v*[LCSM]**$(\mathcal{H}^q)$ will also provide a valid lower bound for optimal objective function value of **[CSM]**. ■

The algorithm starts with a subset $\mathcal{H}^q \subset \mathcal{H}$ of the cuts where $\mathcal{H}^1$ can be empty or chosen a priori while generate the rest as needed. The subset of points $\{M^h_{jt\omega}\}_{\mathcal{H}^q \subset \mathcal{H}}$ are required to obtain the initial subset of cuts and is used to approximate function $R_{ljt\omega}(M_{jt\omega})$. The resulting set is then used to obtain ***v*[LCSM]**$(\mathcal{H}^q)$ which provides a valid lower bound for the original problem **[CSM]** (as shown in **Proposition 1**). We then use this solution to

obtain an upper bound for the CG algorithm. The following proposition is used to obtain an upper bound for the CG algorithm.

Proposition 2 *For any given subset of points* $\{M^h_{jt\omega}\}_{\mathcal{H}^q \subset \mathcal{H}}$, (4.34) *provides the upper bound for the optimal objective function value of* ***[CSM]****.*

$$
\begin{aligned}
UB = \sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt} + \sum_{\omega\in\Omega}\rho_\omega\Bigg(\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt\omega} + \\
\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{bikt\omega} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijkt\omega} + \\
\sum_{k\in\mathcal{K}}p_{bkt}P_{bkt\omega} + \sum_{k\in\mathcal{K}}h_{bkt}H_{bkt\omega} + \sum_{j\in\mathcal{J}}c_0\Bigg(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijkt\omega}}{\sum_{l\in\mathcal{L}}C_{ij}Y_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{bijkt\omega}}\Bigg)\Bigg) + \\
\sum_{k\in\mathcal{K}}\pi_{kt}U_{kt\omega}\Bigg)\Bigg)
\end{aligned}
\tag{4.34}
$$

Proof: Any solution feasible to **[LCSM]**($\mathcal{H}^q$) also provides a feasible solution to **[CSM]** since all the constraints of **[CSM]** are contained by **[LCSM]**($\mathcal{H}^q$). Thus, the objective function value of **[CSM]** evaluated at $(\mathbf{Y}^q, \mathbf{Z}^q, \mathbf{X}^q, \mathbf{P}^q, \mathbf{H}^q, \mathbf{U}^q)$, as shown in equation (4.34), provides an upper bound for the optimal objective value of **[CSM]**. ■

The algorithm continues until the gap between the lower and upper bound falls below a tolerance level ϵ; otherwise, a new set of points $\{M^{h_{new}}_{jt\omega}\}$ are generated using the current solution (shown below) and the process continues.

$$
M^{h_{new}}_{jt\omega} = \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X^q_{bijkt\omega}}{\sum_{l\in\mathcal{L}}C_{lj}Y^q_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X^q_{bijkt\omega}} \tag{4.35}
$$

A pseudo-code of the Constraint Generation algorithm CG is provided in **Algorithm 1**.

Algorithm 1: Constraint Generation Algorithm

Initialize, $q \leftarrow 1$, ϵ, $UB^q \leftarrow +\infty$, $LB^q \leftarrow -\infty$
terminate $\leftarrow$ **false**
Choose an initial set of points: $\{M^h_{jt\omega}\}_{\mathcal{H}^q \subset \mathcal{H}}$
while (*terminate* $=$ **false**) **do**
    Solve **[LCSM]**($\mathcal{H}^q$) to obtain $\boldsymbol{\nu}$**[LCSM]**$(\mathcal{H}^q)$ and $(\mathbf{Y}^q, \mathbf{Z}^q, \mathbf{X}^q, \mathbf{P}^q, \mathbf{H}^q, \mathbf{U}^q)$
    Update the lower bound: $LB^q \leftarrow \boldsymbol{\nu}$**[LCSM]**$(\mathcal{H}^q)$ using (4.33)
    Update the upper bound UB^q using (4.34)
    if $((UB^q - LB^q)/UB^q \leq \epsilon)$ **then**
        terminate $\leftarrow$ **true**
    else
        $M^{h_{new}}_{jt\omega} = \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X^q_{bijkt\omega}}{\sum_{l\in\mathcal{L}} C_{lj} Y^q_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X^q_{bijkt\omega}}$
        $M^{h,q+1}_{jt\omega} = M^{h,q}_{jt\omega} \cup \{M^{h_{new}}_{jt\omega}\}$
    end if
    $q \leftarrow q + 1$
end while

4.3.2 Sample Average Approximation

Obtaining a lower bound by solving problem **[LCSM]**($\mathcal{H}^q$) in the CG algorithm is still considered challenging. Commercial solvers, such as CPLEX fails to solve a moderate size of this problem. To overcome this challenge, Sample Average Approximation (SAA) algorithm is employed to reduce the computational burden to solve problem **[LCSM]**($\mathcal{H}^q$). SAA has previously applied to solve large scale supply chain network flow related problems, such as [116], [102], [16], [103] and others. SAA provides high quality feasible solutions along with the statistical estimation of their optimality gap. Interested readers can review the works from Kleywegt et al. [60] for the proof of convergence properties of SAA and Norkin et al. [87], [86], and Mark et al. [71] for the evaluation of statistical performance of SAA. In SAA, a sample ω^1, ω^2,....., ω^N of N sample scenarios are generated from Ω according to a probability distribution $\mathbb{P}$ and they are solved repeatedly until a

pre-specified tolerance gap is achieved. Therefore, instead of solving the original problem with a large number of scenarios (Ω), the SAA problem is evaluated in each iteration with a smaller number of scenarios. The lower bound of the CG algorithm defined by equation (4.33) subject to constraints (4.20)-(4.32) is now approximated by the following SAA problem:

$$LB = \boldsymbol{v}\textbf{[LCSM]}(\mathcal{H}_N^q) = \sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt} + \frac{1}{N}\sum_{i=1}^{N}\Bigg(\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn} + \sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{biktn} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijktn} + \sum_{k\in\mathcal{K}}p_{bkt}P_{bktn} + \sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Bigg) + \sum_{j\in\mathcal{J}}c_0M_{jtn} + \sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}\Bigg)\Bigg) \tag{4.36}$$

As the sample size (value of N) increases, the optimal solution of **[LCSM]**($\mathcal{H}^q$), i.e., $(\mathbf{Y}^q, \mathbf{Z}^q, \mathbf{X}^q, \mathbf{P}^q, \mathbf{H}^q, \mathbf{U}^q)$ and the optimal objective value $\boldsymbol{v}$**[LCSM]**$(\mathcal{H}^q)$ converges with probability one to an optimal solution of the original problem **[CSM]** [60]. Assuming that the SAA problem is solved within an absolute optimality gap $\delta \geq 0$, we can estimate the sample size N needed to guarantee an ϵ-optimal solution to the true problem with probability at least equal to $(1-\alpha)$ as:

$$N \geq \frac{3\sigma_{max}^2}{(\epsilon-\delta)^2}\Bigg(|\mathcal{L}||\mathcal{J}||\mathcal{T}|(log2) - log\alpha\Bigg) \tag{4.37}$$

where $\epsilon > \delta$, $\alpha \in (0,1)$ and σ_{max}^2 is a maximal variance of certain function differences [60]. Sample size estimation using equation (4.37) is too conservative for practical applications. Hence, selection of a sample size N will be a trade-off between the solution quality

obtained by solving equation (4.36) to the original problem (4.33) and the computational burden needed to solve it. Solutions of the SAA problems with a sample size N provides a statistical lower and upper bounds for the original problem and the process terminates when the gap between the estimators falls below a pre-specified threshold value. The steps involved in solving problem **[LCSM]**($\mathcal{H}^q$) using SAA is given below:

1. Generate M independent feedstock supply scenarios of size N

   i.e., $\{\mathbf{s}_m^1(\omega), \mathbf{s}_m^2(\omega), ..., \mathbf{s}_m^N(\omega)\}$, $\forall m = 1, ..., M$, where $\mathbf{s} = \{s_{bi\omega}, \forall b \in \mathcal{B}, i \in \mathcal{I}, \omega \in \Omega\}$. The lower bound problem for CG algorithm, defined by (4.33) and subject to constraints (4.20)-(4.32), can now be approximated by the following SAA problem:

$$\underset{Y \in \mathbf{Y}}{Minimize}\ \hat{\mathbf{g}}(Y_N^q) = \sum_{t \in \mathcal{T}} \Bigg(\sum_{l \in \mathcal{L}} \sum_{j \in \mathcal{J}} \psi_{ljt} Y_{ljt} + \frac{1}{N} \sum_{i=1}^{N} \Bigg(\sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \xi_{jkt} Z_{jktn} + \tag{4.38}$$

$$\sum_{b \in \mathcal{B}} \Bigg(\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \big(c_{bikt} + \delta_{bit} + \gamma_{bit}\big) X_{biktn} + \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \big(c_{bijkt} + \delta_{bit} + \gamma_{bit}\big) X_{bijktn} +$$

$$\sum_{k \in \mathcal{K}} p_{bkt} P_{bktn} + \sum_{k \in \mathcal{K}} h_{bkt} H_{bktn} \Bigg) + \sum_{j \in \mathcal{J}} c_0 M_{jtn} + \sum_{k \in \mathcal{K}} \pi_{kt} U_{ktn} \Bigg) \Bigg)$$

   Subject to

$$\sum_{k \in \mathcal{K}} X_{biktn} + \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} X_{bijktn} \leq s_{bitn} \qquad \forall b \in \mathcal{B}, i \in \mathcal{I}, t \in \mathcal{T}, n \in N \tag{4.39}$$

$$\sum_{i \in \mathcal{I}} X_{biktn} + \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} X_{bijktn} + (1 - \alpha_b) H_{k,t-1,n} = H_{bktn} + P_{bktn}\ \forall b \in \mathcal{B}, k \in \mathcal{K}, t \in \mathcal{T}, n \in N \tag{4.40}$$

$$\sum_{b \in \mathcal{B}} P_{bktn} + U_{ktn} = b_{kt} \qquad \forall k \in \mathcal{K}, t \in \mathcal{T}, n \in N \tag{4.41}$$

$$\sum_{b\in\mathcal{B}} H_{bktn} \leq h_k^{cap} \qquad \forall k\in\mathcal{K}, t\in\mathcal{T}, n\in N \tag{4.42}$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}} X_{bijktn} \leq v^{cap} Z_{jktn} \qquad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, n\in N \tag{4.43}$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijktn} \leq \sum_{l\in\mathcal{L}} C_{lj} R_{ljtn} \quad \forall j\in\mathcal{J}, t\in\mathcal{T}, n\in N \tag{4.44}$$

$$\sum_{l\in\mathcal{L}} R_{ljtn} \leq \frac{M_{jtn}}{\left(1+M_{jtn}^h\right)^2} + \left(\frac{M_{jtn}^h}{1+M_{jyn}}\right)^2 \qquad \forall j\in\mathcal{J}, t\in\mathcal{T}, n\in N, h\in\mathcal{H} \tag{4.45}$$

$$R_{ljtn} \leq Y_{ljt} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T}, n\in N \tag{4.46}$$

$$\sum_{l\in\mathcal{L}} Y_{ljt} \leq 1 \qquad \forall j\in\mathcal{J}, t\in\mathcal{T} \tag{4.47}$$

$$Z_{jktn} \leq \sum_{l\in\mathcal{L}} \left\lceil \frac{C_{lj}}{C^{cap}} \right\rceil Y_{ljt} \;\forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, n\in N \tag{4.48}$$

$$Y_{ljt} \in \mathbb{B} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T} \tag{4.49}$$

$$Z_{jktn} \in Z^+ \qquad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, n\in N \tag{4.50}$$

$$X_{bijktn}, X_{biktn}, H_{bktn},$$

$$P_{bktn}, U_{ktn}, R_{ljtn}, M_{jtn} \geq 0 \qquad \forall b\in\mathcal{B}, i\in\mathcal{I}, j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, n\in N \tag{4.51}$$

The optimal objective value is denoted by $\mathbf{v}_m^N$ and the optimal solution by $\hat{\mathbf{Y}}_N^m; m = 1, ..., M.$

2. Compute the average of the optimal solutions obtained by solving all SAA problems, $\bar{\mathbf{v}}_M^N$ and variance, $\sigma^2_{\bar{\mathbf{v}}_M^N}$:

$$\bar{\mathbf{v}}_M^N = \frac{1}{M}\sum_{m=1}^{M}\mathbf{v}_N^m$$

$$\sigma^2_{\bar{\mathbf{v}}_M^N} = \frac{1}{(M-1)M}\sum_{m=1}^{M}\left(\mathbf{v}_N^m - \bar{\mathbf{v}}_M^N\right)^2$$

where, $\bar{\mathbf{v}}_M^N$ provides a statistical lower bound on the optimal objective function value for the original problem (4.33) subject to equation (4.20)-(4.32) [87].

3. Pick a feasible first-stage solution obtained from **Step 1** i.e., one of $\hat{\mathbf{Y}}_N^m$ and solve the original problem (4.33) using a reference sample N' as follow:

$$\underset{Y\in\mathbf{Y}}{Minimize}\ \tilde{\mathbf{g}}_{N'}(\tilde{Y}) = \sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}\tilde{Y}_{ljt} + \frac{1}{N'}\sum_{i=1}^{N'}\Bigg(\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn} + \tag{4.52}$$

$$\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{biktn} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijktn} +$$

$$\sum_{k\in\mathcal{K}}p_{bkt}P_{bktn} + \sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Bigg) + \sum_{j\in\mathcal{J}}c_0M_{jtn} + \sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}\Bigg)\Bigg)$$

The estimator $\tilde{\mathbf{g}}_{N'}(\tilde{Y})$ serves as an upper bound for the problem **[LCSM]($\mathcal{H}^q$)** and will be updated in each iteration if the value obtained is less than the value of the previous iteration. N' is another set of samples generated independently for the feedstock supply scenarios, i.e., $\{\mathbf{s}^1(\omega), \mathbf{s}^2(\omega), ..., \mathbf{s}^{N'}(\omega)\}$, $\forall n = 1, ..., N'$. Typically, the sample size N' is chosen much larger than the sample size N used in the SAA problems i.e., $N' \gg N$. The variance of $\tilde{\mathbf{g}}_{N'}(\tilde{Y})$ is estimated as follow:

$$\sigma^2_{N'}(\tilde{Y}) = \frac{1}{(N'-1)N'}\sum_{n=1}^{N'}\Big\{\sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}\tilde{Y}_{ljt} + \mathbb{Q}(\mathbf{Y}, n)\Bigg) - \tilde{\mathbf{g}}_{N'}(\tilde{Y})\Big\}^2$$

where $\mathbb{Q}(\mathbf{Y}, n)$ represents the solution of the second-stage problem.

4. Compute the optimality gap ($gap_{N,M,N'}(\tilde{Y})$) and its variance (σ^2_{gap}) using the estimators calculated in **Steps 2** and **3**.

$$\begin{aligned} gap_{N,M,N'}(\tilde{Y}) &= \tilde{\mathbf{g}}_{N'}(\tilde{Y}) - \bar{\mathbf{v}}^N_M \\ \sigma^2_{gap} &= \sigma^2_{N'}(\tilde{Y}) + \sigma^2_{\bar{\mathbf{v}}^N_M} \end{aligned}$$

The confidence interval for the optimality gap is then calculated as follows:

$$\tilde{\mathbf{g}}_{N'}(\tilde{Y}) - \bar{\mathbf{v}}^N_M + z_\alpha \left\{ \sigma^2_{N'}(\tilde{Y}) + \sigma^2_{\bar{\mathbf{v}}^N_M} \right\}^{1/2}$$

with z_α:= $\Phi^{-1}(1-\alpha)$, where $\Phi(z)$ is the cumulative distribution function of the standard normal distribution.

4.3.3 Progressive Hedging

Step 1 in SAA algorithm involves in solving a two-stage stochastic mixed-integer linear programming model of N scenarios. The problem is still considered challenging from solution point of view; thus, it motivates us to solve each subproblems of the SAA problem using a Progressive Hedging Algorithm (PHA) [99]. The PHA proceeds by applying a scenario decomposition technique based on the augmented Lagrangian relaxation scheme to solve a number of individual scenario subproblems and finally aggregating the individual scenario solutions. Progressive Hedging algorithm has successfully applied to a number of different application areas, such as financial planning [84], fisheries management [47],

surgery planning [44], hydrothermal operation planning [101], [14], and others. For a detail understanding about the method, readers can review the studies conducted by Wallace and Helgason [119] and Watson and Woodruff [123].

Note that constraints (4.46) and (4.48) in $\hat{\mathbf{g}}(Y_N^q)$ (shown in **Step 1** on SAA algorithm) link the first-stage decisions with the second-stage decision variables. These constraints will not allow problem $\hat{\mathbf{g}}(Y_N^q)$ separable by scenarios. To remedy this problem, we define a new decision variable $\{Y_{ljtn}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T},n\in N} \in \{0,1\}$ that creates a copy of the first-stage decision variable for scenario $n \in N$. Problem $\hat{\mathbf{g}}(Y_N^q)$ can now be rewritten as follows:

$$\underset{\mathbf{Y,Z,X,P,H,M,U}}{Minimize}\ \frac{1}{N}\sum_{n=1}^{N}\sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljtn}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn}+$$

$$\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{biktn}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijktn}+$$

$$\sum_{k\in\mathcal{K}}p_{bkt}P_{bktn}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Bigg)+\sum_{j\in\mathcal{J}}c_0M_{jtn}+\sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}\Bigg)$$

Subject to: (4.39)-(4.45), (4.50), (4.51), and

$$R_{ljtn} \leq Y_{ljtn} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T}, n\in\mathcal{N} \tag{4.53}$$

$$\sum_{l\in\mathcal{L}}Y_{ljtn} \leq 1 \qquad \forall j\in\mathcal{J}, t\in\mathcal{T}, n\in\mathcal{N} \tag{4.54}$$

$$Z_{jktn} \leq \sum_{l\in\mathcal{L}}\left\lceil\frac{C_{lj}}{c^{cap}}\right\rceil Y_{ljtn} \qquad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T}, n\in N \tag{4.55}$$

$$Y_{ljtn} = Y_{ljtm} \qquad \forall n,m\in N, n\neq m \tag{4.56}$$

$$Y_{ljtn} \in \mathbb{B} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T}, n\in N \tag{4.57}$$

Constraints (4.56) are referred to as *nonanticipativity* constraints. These constraints force the model not to be separable by scenarios. We need to rewrite these *nonanticipa-*

tivity constraints in order to make the model separable by scenarios and applicable to use Lagrangian relaxation technique. Let $\{\bar{Y}_{ljt}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}} \in \{0,1\}$ be the "overall design vector". The following constraints are equivalent to (4.56):

$$Y_{ljtn} = \bar{Y}_{ljt} \qquad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, t \in \mathcal{T}, n \in N \tag{4.58}$$

$$\bar{Y}_{ljt} \in \mathbb{B} \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T} \tag{4.59}$$

We will now follow the decomposition scheme proposed by Rockafellar and Wets [99] to relax constraints (4.58) using an augmented Lagrangian strategy and obtain the following objective function:

$$\begin{aligned}
&\underset{\mathbf{Y,Z,X,P,H,M,U}}{Minimize}\ \frac{1}{N}\sum_{n=1}^{N}\sum_{t\in\mathcal{T}}\Bigg(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljtn}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn}\\
&+\sum_{b\in\mathcal{B}}\Bigg(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{biktn}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijktn}\\
&+\sum_{k\in\mathcal{K}}p_{bkt}P^{q}_{bktn}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Bigg)+\sum_{j\in\mathcal{J}}c_0M_{jtn}+\sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}\\
&+\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\lambda}_{ljtn}(Y_{ljtn}-\bar{Y}_{ljt})+\frac{1}{2}\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\phi}(Y_{ljtn}-\bar{Y}_{ljt})^2\Bigg)
\end{aligned}$$

where $\{\boldsymbol{\lambda}_{ljtn}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T},n\in N}$ defines the Lagrangian multipliers for the relaxed constraints and $\boldsymbol{\phi}$ defines a penalty ratio. Given the binary requirements of variables $\{Y_{ljtn}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T},n\in N}$ and $\{\bar{Y}_{ljt}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ the quadratic term $\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\phi}(Y_{ljtn}-\bar{Y}_{ljt})^2$ shown in the above objective function can be reduced as follows:

$$\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\phi}(Y_{ljtn}-\bar{Y}_{ljt})^2 = \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\Big(\boldsymbol{\phi}(Y_{ljtn})^2-2\boldsymbol{\phi}Y_{ljtn}\bar{Y}_{ljt}+\boldsymbol{\phi}(\bar{Y}_{ljt})^2\Big)$$
$$= \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\Big(\boldsymbol{\phi}Y_{ljtn}-2\boldsymbol{\phi}Y_{ljtn}\bar{Y}_{ljt}+\boldsymbol{\phi}\bar{Y}_{ljt}\Big)$$

The objective function, therefore, can be reduced as follows:

$$\underset{\mathbf{Y,Z,X,P,H,M,U}}{Minimize}\ \frac{1}{N}\sum_{n=1}^{N}\sum_{t\in\mathcal{T}}\Big(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}(\psi_{ljt}+\boldsymbol{\lambda}_{ljtn}-\boldsymbol{\phi}\bar{Y}_{ljt}$$
$$+\frac{\boldsymbol{\phi}}{2})Y_{ljtn}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn}+\sum_{b\in\mathcal{B}}\Big(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{biktn}$$
$$+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijktn}+\sum_{k\in\mathcal{K}}p_{bkt}P_{bktn}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Big)$$
$$+\sum_{j\in\mathcal{J}}c_0M_{jtn}+\sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}-\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\lambda}_{ljtn}\bar{Y}_{ljt}-\frac{1}{2}\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\boldsymbol{\phi}\bar{Y}_{ljt}\Big)$$

When the value of the overall plan $\{\bar{Y}^q_{ljt}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ is fixed, the last two terms of the above objective function becomes constant and thus can be decomposed into a series of deterministic problem of $n \in N$ scenarios. The overall problem can be formulated for each scenario $n \in N$ as follows:

$$\textbf{[LCSM(PHA)]}\ \underset{\mathbf{Y,Z,X,P,H,M,U}}{Minimize}\ \sum_{t\in\mathcal{T}}\Big(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}(\psi_{ljt}+\boldsymbol{\lambda}_{ljtn}-\boldsymbol{\phi}\bar{Y}_{ljt}+\frac{\boldsymbol{\phi}}{2})Y_{ljtn}$$
$$+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jktn}+\sum_{b\in\mathcal{B}}\Big(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\big(c_{bikt}+\delta_{bit}+\gamma_{bit}\big)X_{biktn}$$
$$+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\big(c_{bijkt}+\delta_{bit}+\gamma_{bit}\big)X_{bijktn}+\sum_{k\in\mathcal{K}}p_{bkt}P_{bktn}+\sum_{k\in\mathcal{K}}h_{bkt}H_{bktn}\Big)$$
$$+\sum_{j\in\mathcal{J}}c_0M_{jtn}+\sum_{k\in\mathcal{K}}\pi_{kt}U_{ktn}\Big)$$

Subject to

$$\sum_{k\in\mathcal{K}} X_{biktn} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} X_{bijktn} \leq s_{bitn} \qquad \forall b\in\mathcal{B}, i\in\mathcal{I}, t\in\mathcal{T} \quad (4.60)$$

$$\sum_{i\in\mathcal{I}} X_{biktn} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} X_{bijktn} + (1-\alpha_b)H_{k,t-1,n} = H_{bktn} + P_{bktn} \qquad \forall b\in\mathcal{B}, k\in\mathcal{K}, t\in\mathcal{T} \quad (4.61)$$

$$\sum_{b\in\mathcal{B}} P_{bktn} + U_{ktn} = b_{kt} \qquad \forall k\in\mathcal{K}, t\in\mathcal{T} \quad (4.62)$$

$$\sum_{b\in\mathcal{B}} H_{bktn} \leq h_k^{cap} \qquad \forall k\in\mathcal{K}, t\in\mathcal{T} \quad (4.63)$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}} X_{bijktn} \leq v^{cap} Z_{jktn} \qquad \forall (j,k)\in\mathcal{A}_2, t\in\mathcal{T} \quad (4.64)$$

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{bijktn} \leq \sum_{l\in\mathcal{L}} C_{lj} R_{ljtn} \qquad \forall j\in\mathcal{J}, t\in\mathcal{T} \quad (4.65)$$

$$\sum_{l\in\mathcal{L}} R_{ljtn} \leq \frac{M_{jtn}}{\left(1+M_{jtn}^h\right)^2} + \left(\frac{M_{jtn}^h}{1+M_{jtn}}\right)^2 \qquad \forall j\in\mathcal{J}, t\in\mathcal{T}, h\in\mathcal{H} \quad (4.66)$$

$$R_{ljtn} \leq Y_{ljtn} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T} \quad (4.67)$$

$$\sum_{l\in\mathcal{L}} Y_{ljtn} \leq 1 \qquad \forall j\in\mathcal{J}, t\in\mathcal{T} \quad (4.68)$$

$$Z_{jktn} \leq \sum_{l\in\mathcal{L}} \left\lceil \frac{C_{lj}}{c^{cap}} \right\rceil Y_{ljtn} \qquad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T} \quad (4.69)$$

$$Y_{ljtn} \in \mathbb{B} \qquad \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T} \quad (4.70)$$

$$Z_{jktn} \in Z^+ \qquad \forall j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T} \quad (4.71)$$

$$X_{bijktn}, X_{biktn}, H_{bktn}, P_{bktn}, U_{ktn}, R_{ljtn}, M_{jtn} \geq 0 \qquad \forall b\in\mathcal{B}, i\in\mathcal{I}, j\in\mathcal{J}, k\in\mathcal{K}, t\in\mathcal{T} \quad (4.72)$$

Let r denote the index for the iteration number for the PHA and $\{\boldsymbol{\lambda}_{ljtn}^r\}_{\forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T}, n\in N}$ is the lagrangian multipliers and $\boldsymbol{\phi}^r$ is the penalty parameter at

iteration r. The general idea of the basic PHA is to solve N deterministic **[LCSM(PHA)]** problem and obtain the consensus parameter $\{\bar{Y}^r_{ljt}\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$. If the gap between the binary variable Y^r_{ljtn} and the consensus parameter $\bar{Y}^r_{ljt}$ falls below a threshold value ϵ (i.e., $\epsilon = 0.001$) for each $l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$ then the algorithm terminates; otherwise, the values of $\boldsymbol{\lambda}^r_{ljtn}$ and $\boldsymbol{\phi}^r$ are updated using equations (4.73) and (4.74) and the process continues.

$$\boldsymbol{\lambda}^r_{ljtn} \leftarrow \boldsymbol{\lambda}^{r-1}_{ljtn} + \boldsymbol{\phi}^{r-1}_{ljt}(Y^r_{ljtn} - \bar{Y}^{r-1}_{ljt}) \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T} \quad (4.73)$$

$$\boldsymbol{\phi}^r \leftarrow \alpha\boldsymbol{\phi}^{r-1} \quad (4.74)$$

where $\boldsymbol{\lambda}^0_{ljtn}$ is set to zero for each scenario $n \in N$; $\boldsymbol{\phi}^0$ is set to a fixed positive value which ensures that $\boldsymbol{\phi}^r \to \infty$ as the number of iteration r increases and the constant parameter $\alpha > 1$. A pseudo-code of the basic PHA is provided in **Algorithm 2**.

Termination Criteria: The Progressive Hedging algorithm terminates when one of the following condition is satisfied:

- $\frac{1}{N}\sum_{n=1}^{N}\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}|Y^r_{ljtn} - \bar{Y}^r_{ljt}| \leq \epsilon$; where ϵ is a pre-specified tolerance gap
- 10 consecutive non-improvement iterations
- Maximum iteration limit is reached (i.e., $iter^{max}$ = 100)
- Maximum time limit is reached (i.e., $time^{max}$ = 10,800 CPU seconds)

Algorithm 2: Progressive Hedging Algorithm

Initialize, $r \leftarrow 1$, ϵ, $\{\boldsymbol{\lambda}^r_{ljtn}\}_{\forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}, n \in N} \leftarrow 0$, $\boldsymbol{\phi}^r \leftarrow \boldsymbol{\phi}^0$
terminate $\leftarrow$ **false**
while (*terminate* $=$ **false**) **do**
    for $n = 1$ to N
        Solve **[LCSM(PHA)]** and obtain $\{Y^r_{ljtn}\}_{\forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}, n \in N}$
    end for
    Calculate the consensus parameter:
        $\bar{Y}^r_{ljt} \leftarrow \frac{1}{N} \sum_{n=1}^{N} Y^r_{ljtn}; \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$
    if $(r > 1)$ **then**
        Update the lagrangian parameter:
            $\boldsymbol{\lambda}^r_{ljtn} \leftarrow \boldsymbol{\lambda}^{r-1}_{ljtn} + \boldsymbol{\phi}^{r-1}_{ljt}(Y^r_{ljtn} - \bar{Y}^{r-1}_{ljt}); \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$
        Update the penalty parameter:
            $\boldsymbol{\phi}^r \leftarrow \alpha \boldsymbol{\phi}^{r-1}$ and $\alpha > 1$
    end if
    if $((Y^r_{ljtn} - \bar{Y}^{r-1}_{ljt})_{\forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}} \leq \epsilon)$ **then**
        terminate $\leftarrow$ **true**
    end if
    $r \leftarrow r + 1$
end while

4.3.4 Enhanced Progressive Hedging Algorithm

Our initial computational experimentation with PHA to solve **[LCSM(PHA)]** for a sufficiently large network size problem exposes its inability to converge within a reasonable amount of time. This motivates us to explore additional enhancement techniques to improve the convergence and stability of the PHA algorithm. The following subsection discusses some PHA enhancement techniques to solve problem **[LCSM(PHA)]** efficiently in a reasonable amount of time.

4.3.4.1 Penalty Parameter Updating

A number of studies, such as Chen and Fan [17], Huang et al. [50], showed that setting of $\boldsymbol{\phi}$ strongly influences the performance of the PHA algorithm. The authors observed that

when the value of ϕ set too high then the algorithm converges fast to a suboptimal solution. Conversely, if we set a conservative value for ϕ, then the algorithm converges slowly to a near optimal solution. To tackle this challenge, this study uses a method proposed by Hvattum and Lokketangen [53] to dynamically adjust the value of ϕ over iterations based on the computational performance obtained from prior iterations of the PHA algorithm. Let $\boldsymbol{\Delta}_1^r$ and $\boldsymbol{\Delta}_2^r$ be indicators of the convergence rates in the dual space and in the primal space, respectively. Thus, the penalty value can be updated as follows:

$$\boldsymbol{\Delta}_1^r = \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}(Y_{ljt}^r - \bar{Y}_{ljt}^r)^2 \tag{4.75}$$

$$\boldsymbol{\Delta}_2^r = \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}(\bar{Y}_{ljt}^r - \bar{Y}_{ljt}^{r-1})^2 \tag{4.76}$$

$$\boldsymbol{\phi}^r = \begin{cases} \varphi\boldsymbol{\phi}^{r-1} & \text{if } \boldsymbol{\Delta}_1^r - \boldsymbol{\Delta}_1^{r-1} > 0 \\ \frac{1}{\varphi}\boldsymbol{\phi}^{r-1} & \text{else if } \boldsymbol{\Delta}_2^r - \boldsymbol{\Delta}_2^{r-1} > 0 \end{cases} \tag{4.77}$$

where φ is a constant parameter which value is set to $\varphi > 1$.

4.3.4.2 Heuristic Strategy

We will now use two heuristic strategies that modify the fixed cost of using the multi-modal facilities in problem **[LCSM(PHA)]** and enhances the PHA algorithm. The first strategy is termed as *global heuristic* [19] since this strategy modifies the fixed cost of using a multi-modal facility at the end of each iteration. Note that problem **[LCSM(PHA)]** is composed of a series of N deterministic sub-problems. In **Algorithm 2**, at the end of each iteration r one can obtain the values of the consensus parameter $\{\bar{Y}_{ljt}^r\}_{\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$

which provide an indication of how many times a multi-modal facility was used in the previous iterations. A higher value of $\bar{Y}^r_{ljt}$ translates the fact that the multi-modal facility j of capacity l at time period t was used in most of the previous iterations. Conversely, a lower value of $\bar{Y}^r_{ljt}$ indicates that the multi-modal facility j of capacity l at time period t was not a favorable decision in most of the previous iterations. Let $\overline{c}$ and $\underline{c}$ be the two parameters that define an upper and lower threshold value. Therefore, if the value of $\bar{Y}^r_{ljt}$ is greater that the threshold value $\overline{c}$, then lowering the fixed cost of using the multi-modal facility will attract the subproblems to use the facility in the coming iterations. Similarly, if the value of $\bar{Y}^r_{ljt}$ is lower that the threshold value $\underline{c}$, then increasing the fixed cost of using the multi-modal facility will avoid the subproblems use the facility in the coming iterations. This will fix the decisions of using few multi-modal facilities to either one or zero and thus will help to reduce the size of the problem. The adjustment strategy is shown below:

$$\psi^r_{ljt} = \begin{cases} \beta\psi^{r-1}_{ljt} & \text{if } \bar{Y}^{r-1}_{ljt} < \underline{c} \\ \frac{1}{\beta}\psi^{r-1}_{ljt} & \text{if } \bar{Y}^{r-1}_{ljt} > \overline{c} \\ \psi^{r-1}_{ljt} & \text{Otherwise} \end{cases} \tag{4.78}$$

where ψ^r_{ljt} represents the modified fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$ and at iteration r; $\underline{c}$ and $\overline{c}$ are the two constant parameters whose values are set to $0 < \underline{c} < 0.3$ and $0.7 < \overline{c} < 1$; and β is a constant parameter whose value is set to $\beta > 1$.

The above *global heuristic* strategy can be enhanced even further by modifying the fixed cost of using a multi-modal facility within the scenario level. This strategy is termed as *local heuristics* [19] since the modification of the fixed cost only impacts the subproblem of the current scenario at a particular iteration. In scenario $n \in N$ and iteration r, if the gap between Y^r_{ljtn} and $\bar{Y}^r_{ljt}$ is sufficiently large then we modify the fixed cost of using the multi-modal facilities using the following equation. The local adjustment strategy applied to **Algorithm 2** is shown below:

$$\psi^r_{ljtn} = \begin{cases} \beta\psi^r_{ljt} & \text{if } |Y^{r-1}_{ljtn} - \bar{Y}^r_{ljt}| \geq c^{far} \text{ and } Y^{r-1}_{ljtn} = 1 \\ \frac{1}{\beta}\psi^r_{ljt} & \text{if } |Y^{r-1}_{ljtn} - \bar{Y}^r_{ljt}| \geq c^{far} \text{ and } Y^{r-1}_{ljtn} = 0 \\ \psi^r_{ljt} & \text{Otherwise} \end{cases} \tag{4.79}$$

where ψ^r_{ljtn} represents the modified fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$ under scenario $n \in N$ and at iteration r; c^{far} is a threshold at which point a local adjustment to the fixed cost of using a multi-modal facility is applied and is set to $0.5 < c^{far} < 1$; and β is a constant parameter whose value is set to $\beta > 1$.

4.3.4.3 Rolling Horizon

Problem **[LCSM(PHA)]** is a deterministic, multi-time period facility location problem. It is a special case of a capacitated facility location problem which is known to be an $\mathcal{NP}$-hard problem [69]. Therefore, in this section we solve problem **[LCSM(PHA)]** using a Rolling Horizon **[RH]** heuristic approach proposed by Balasubramanian and Grossman

[7] and Kostina et al. [62]. This approach decomposes multi-time period **[LCSM(PHA)]** problem into a series of small subproblems with few consecutive time periods and the algorithm terminates when all the subproblems are investigated. A pseudo-code of the **[RH]** algorithm is shown in **Algorithm 3**

Let t_0^r denote the starting time period of subproblem r. Let $\mathbf{M}^r$ denote the number of time periods comprised in subproblem r. We can set a fixed or variable size of $\mathbf{M}^r$ for each subproblem. Each approximate subproblem of the Rolling horizon algorithm is denoted by **[LCSM(PHA(r))]**. Now, for each scenario $n \in N$ the approximate subproblems are solved by setting the variables as: (i) $\{Y_{ljtn}\}_{\forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}} \in \{0,1\}$ and $\{Z_{jktn}\}_{\forall j \in \mathcal{J}, k \in \mathcal{K}, t \in \mathcal{T}} \in \mathbb{Z}^+$ for $t_0^r \leq t \leq t_0^r + \mathbf{M}^r$, (ii) $0 \leq Y_{ljtn} \leq 1$ and $Z_{jktn} \in \mathbb{R}^+$ for $t > t_0^r + \mathbf{M}^r$. Once a subproblem is solved, we fix the values of $Y_{ljtn}^r = Y_{ljtn}^{r-1}; \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$ and $Z_{jktn}^r = Z_{jktn}^{r-1}; \forall j \in \mathcal{J}, k \in \mathcal{K}, t \in \mathcal{T}$ for $t < t_0^r$ and update the step size r. The process terminates when all the subproblems are solved. **Figure 4.4** shows an example of using the rolling horizon approach to solve a three time period problem.

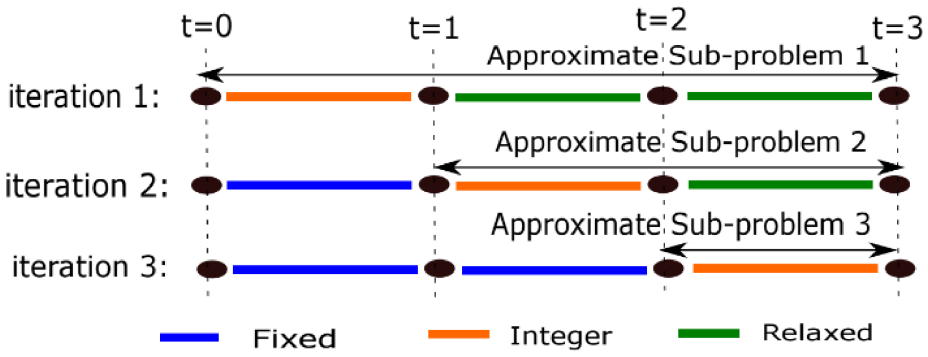


Figure 4.4

Application of a rolling horizon strategy for a three time period problem

Algorithm 3: Rolling Horizon (RH) Heuristic

$r \leftarrow 1$, $t_0^r = 0$, M^r, *terminate* $\leftarrow$ **false**
while (*terminate* $=$ **false**) **do**
  Set:
    $Y_{ljtn} \in \{0, 1\}$ and $Z_{jktn} \in \mathbb{Z}^+$ for $t_0^r \leq t \leq t_0^r + M^r$
    $0 \leq Y_{ljtn} \leq 1$ and $Z_{jktn} \in \mathbb{R}^+$ for $t > t_0^r + M^r$
  Solve the approximate sub-problem **[LCSM(PHA)]** using CPLEX
  if($t_0 > |\mathcal{T}|$) **then**
    stop $\leftarrow$ **true**
  else
    Fixing the values of Y_{ljtn} and Z_{jktn} for $t < t_0^r$
  end if
  $r \leftarrow r + 1$
end while

4.4 Computational Study and Managerial Insights

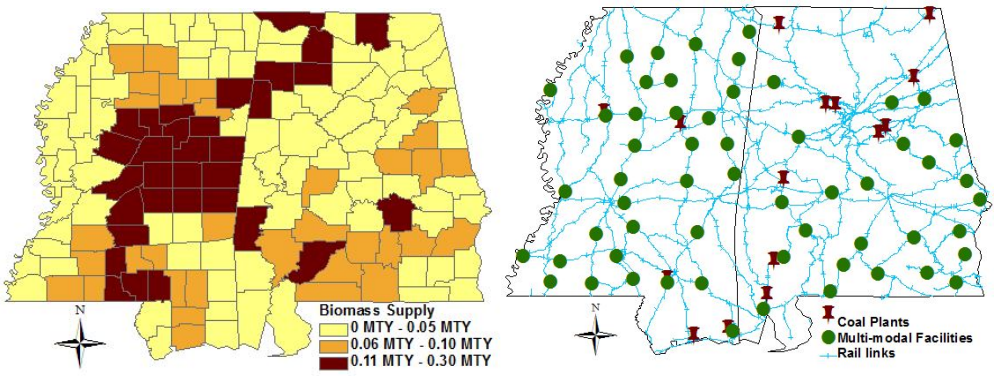
We develop a case study in order to evaluate the impact of congestion in supply chain network performance under feedstock supply uncertainty. We use the states of Mississippi and Alabama as a testing ground for this analysis. All the algorithms are coded in GAMS 24.2.1 [39] and executed on a desktop computer with Intel Core i7 3.50 GHz processor and 32.0 GB RAM. The optimization solver used is ILOG CPLEX 12.6.

4.4.1 Data Description

4.4.1.1 Biomass Supply

Corn stover and forest residues are the two major feedstock sources available in Mississippi and Alabama. Forest residues are available all year around except three months during the winter (December to February) whereas corn stover is harvested only during the fall season (from September until November). The historical biomass availability data was

adopted from the Knowledge Discovery Framework (KDF) database of the United States Department of Energy [10]. This data was further processed by Idaho National Lab (INL) to calculate the amount of densified biomass available in Mississippi and Alabama. Table 4.1 summarizes the key feedstock parameters used in this study. In total 99 suppliers from Mississippi and Alabama are considered whose geographic distribution are shown in Figure 4.5a. The counties that produce more than 10,000 tons of biomass each year are only considered here.



(a) Biomass distribution (b) Network representation

Figure 4.5

Biomass distribution and facility locations

4.4.1.2 Biomass Demand

The data about coal-fired power plant locations, nameplate capacity, types of coal used, efficiency and operating hours of the plants, and annual heat input rates are obtained from the National Energy Technology Laboratory [110]. We have considered in total 15 coal plants which are distributed in Mississippi and Alabama and their geographic distribution

Table 4.1

Key feedstock parameters

Feedstock type	Availability (MTY)	Deterioration rate[b] (per season)	Storage cost[a] ($/dry ton)	Procurement cost[b] ($/dry ton)
Corn-stover	2.45	10%	8.0	35.0
Forest residues	1.91	12%	2.0	30.0

[a]Obtained from Xie et al. [128]
[b]Obtained from Parker et al. [88]

are shown in Figure 4.5b. We assume that the power plants will replace 6% of coal's demand with densified biomass to produce renewable electricity. This results a demand of 5.6 million tons per year (MTY) of densified biomass that is required to be satisfied by the biomass supply chain network.

4.4.1.3 Investment Costs

This study considers a total of 57 potential locations for rail ramps located in Mississippi and Alabama (shown in Figure 4.5b). The annualized fixed cost for a rail ramp of capacity 1.05 MTY that sends a single railcar is set to $54,949/year [70]. We consider four different facility capacities (l = 0.5 MTY, 0.75 MTY, 1.0 MTY, and 1.25 MTY). These costs are estimated based on a lifetime of 30 years and a discount factor of 10% is assumed.

4.4.1.4 Transportation cost

In this study we assume that trucks are used to transport biomass from a feedstock supplier site $i \in \mathcal{I}$ to a multi-modal facility $j \in \mathcal{J}$. Trucks can further be used to transport biomass directly from a feedstock supplier site $i \in \mathcal{I}$ to a coal plant $k \in \mathcal{K}$. Table 4.2

shows the major cost components used to calculate the unit truck transportation cost and are obtained from the study of Parker et al. [88]. This study further assumes that railcar can be used to transport biomass from a multi-modal facility $j \in \mathcal{J}$ to a coal plant $k \in \mathcal{K}$. The fixed and variable cost component for a single rail car that carries 100 tons of grains is set equal to \$2,248 and \$1.12/ton/mile traveled [42]. For all modes of transportation, we used Arc GIS Desktop 10 to create a transportation network for identifying the shortest path between each source to destination pairs. The network includes actual railways as well as local, rural, urban roads, and major highways for the states of Mississippi and Alabama.

Table 4.2

Truck transportation cost components [88]

Item	Value	Unit
Loading/unloading	5.0	\$/wet ton
Time dependent	29.0	\$/hr/truckload
Distance dependent	1.20	\$/mile/truckload
Truck capacity	25	wet tons/truckload
Average travel speed	40	miles/hour

4.4.2 Experimental Results

4.4.2.1 Impact of congestion cost on system performance

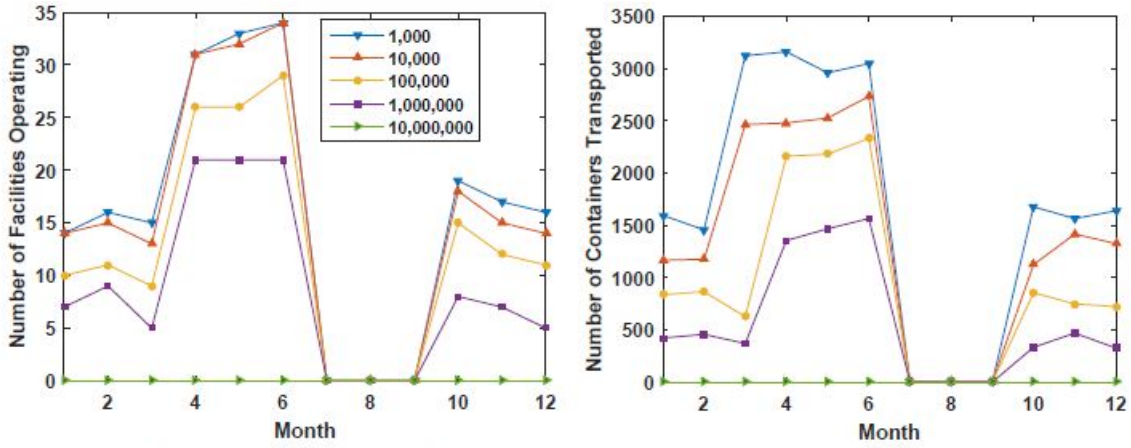
The first set of experiment studies the impact of congestion cost on biomass co-firing supply network decisions. Figure 4.6 provides a relationship between different decision variables under different c_0 values. Clearly, increasing the c_0 values for a given multi-modal facility $j \in \mathcal{J}$ decreases the number of multi-modal facilities $|\mathbf{Y}|$ and thus the number of containers transported between the multi-modal facilities $|\mathbf{Z}|$ by the biomass

supply chain network. This relationship is reflected in Figures **??** and **??**. Furthermore, it is observed that after a certain value of c_0 all the lines become flat. This is the critical point after which no multi-modal facilities $|\mathbf{Y}|$ and thus no containers $|\mathbf{Z}|$ transported between the multi-modal facilities are used to ship biomass in the supply chain network. This in turn increases the amount of biomass shipped via highways (i.e. trucks), as shown in Figure **??**.

We now present the optimal design for biomass co-firing supply chain network while considering a low (e.g., c_0 = \$10,000) and a high congestion cost (e.g., c_0 = \$100,000) into account. Figure 4.7 shows the optimal multi-modal facility deployment under this two experiments. It is observed that in total 34 multi-modal facilities are selected for the case when a lower congestion cost are taken into account (shown in Figure 4.7a). Among the 34 multi-modal facilities 24 are selected with a capacity of 0.5 MTY, 5 with 0.75 MTY, and 5 with 1.25 MTY. On the other hand, we observe a total of 21 multi-modal facilities are selected for the case when a higher congestion cost are taken into account (shown in Figure 4.7b). Among the 21 multi-modal facilities 4 are selected with a capacity of 0.5 MTY, 6 with 0.75 MTY, and 11 with 1.25 MTY. Clearly, as the congestion cost c_0 increases, the number of multi-modal facilities $|\mathbf{Y}|$ decreases but the tendency to select the number of facilities with higher capacity increases. In summary, we can conclude that the congestion cost has a direct impact in the size and selection of location for the multi-modal facilities.

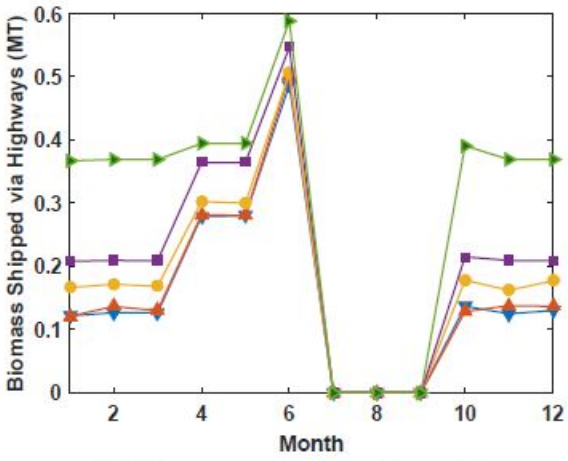
4.4.2.2 Impact of supply changes on system performance

The second set of experiment (shown in Figure 4.8) shows the impact of biomass supply changes on biomass co-firing supply chain network performance. There are a number



(a) Facility operating vs. time periods

(b) Containers transported vs. time periods

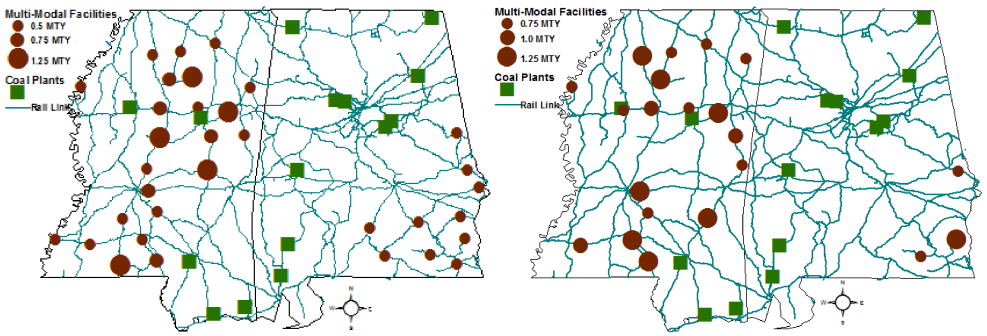


(c) Highway transportation vs. time periods

Figure 4.6

Impact of different congestion cost on system performance

of factors which impacts the biomass availability in a given region, such as, unexpected natural catastrophe, new initiative from government to cultivate (i.e., promoting farmers to cultivate more biomass) or any other related factors [73]. We test model **[LCSM]** by varying the mean biomass supply $\overline{s}_{bit}, \forall b \in \mathcal{B}, i \in \mathcal{I}, t \in \mathcal{T}$ between -20% to +20% and observe that the number of multi-modal facilities and the containers transported between the facilities are significantly impacted by the changes in biomass supply. To run these exper-



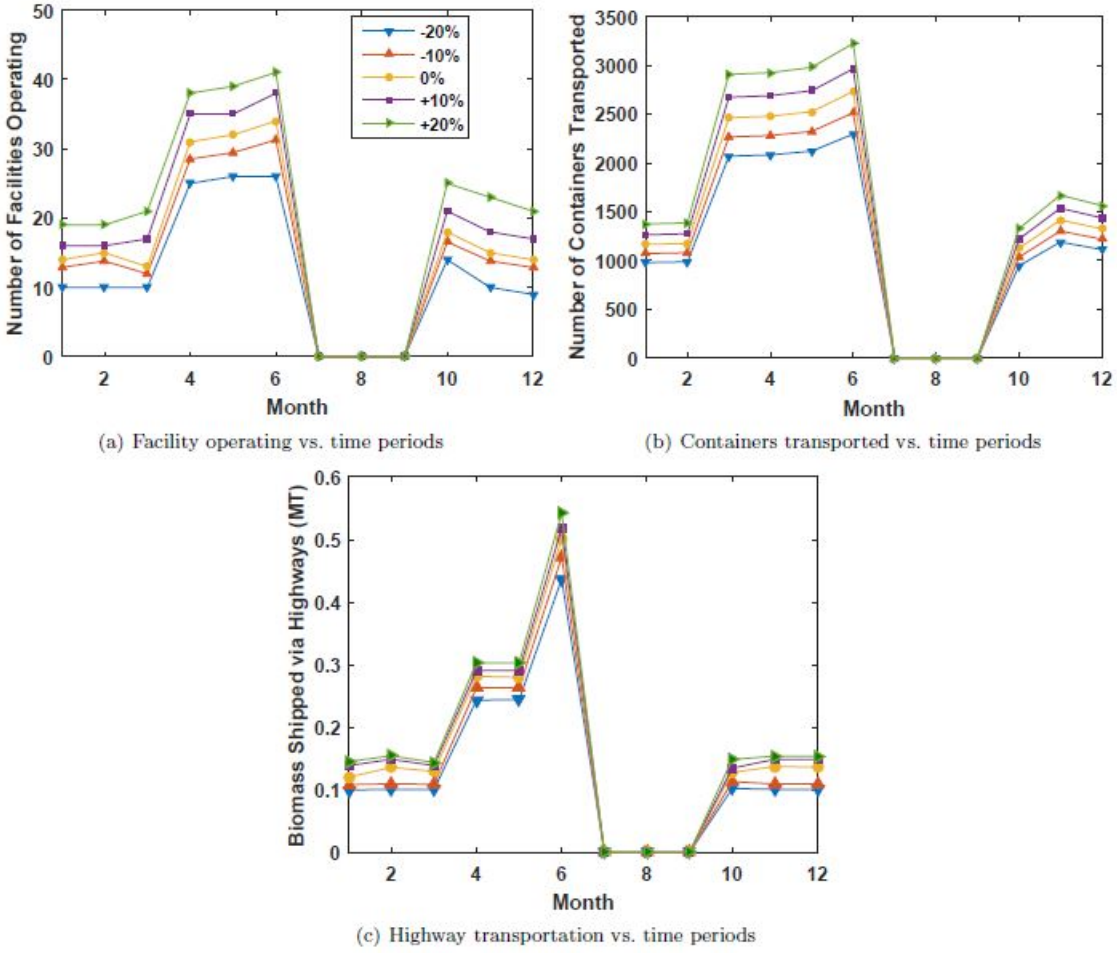
(a) Congestion cost, c_0 = \$10,000 (b) Congestion cost, c_0 = \$100,000

Figure 4.7

Network representation under different congestion costs

iments, we kept the congestion cost fixed for all the instances (i.e., $c_0 = \$10,000$). Figure **??** indicates that if the mean biomass supply $\overline{s}_{bit}$ increases by 20% then model **[LCSM]** selects an additional 32.2% multi-modal facilities $|\mathbf{Y}|$ to cope with this high biomass supply changes. Note that this number would change to 38.1% if the congestion costs were ignored by the model. We observe the same trend among the number of containers $|\mathbf{Z}|$ transported between the multi-model facilities. Results show that a 20% increase in mean biomass supply $\overline{s}_{bit}$ increases $|\mathbf{Z}|$ by 18.04% if the congestion costs were taken into account (shown in Figure **??**) while the number would reach to 25.52% if the congestion costs were ignored. We further observe a 17.64% increase in highway transportation $|\mathbf{X}|$ across all possible scenarios if the mean biomass supply $\overline{s}_{bit}$ would increase by 20%. On the other hand, when we assume that the mean biomass supply $\overline{s}_{bit}$ is decreased by 20%, the number of multi-modal facilities $|\mathbf{Y}|$, the number of containers $|\mathbf{Z}|$ transported between the multi-modal facilities, and the shipment through highways $|\mathbf{X}|$ would decrease by 24.73%,

16.12%, and 12.16%, respectively. This clearly indicates that the changes in mean biomass supply $\overline{s}_{bit}$ will highly impact the biomass supply chain network decisions.



(a) Facility operating vs. time periods

(b) Containers transported vs. time periods

(c) Highway transportation vs. time periods

Figure 4.8

Impact of supply changes on system performance

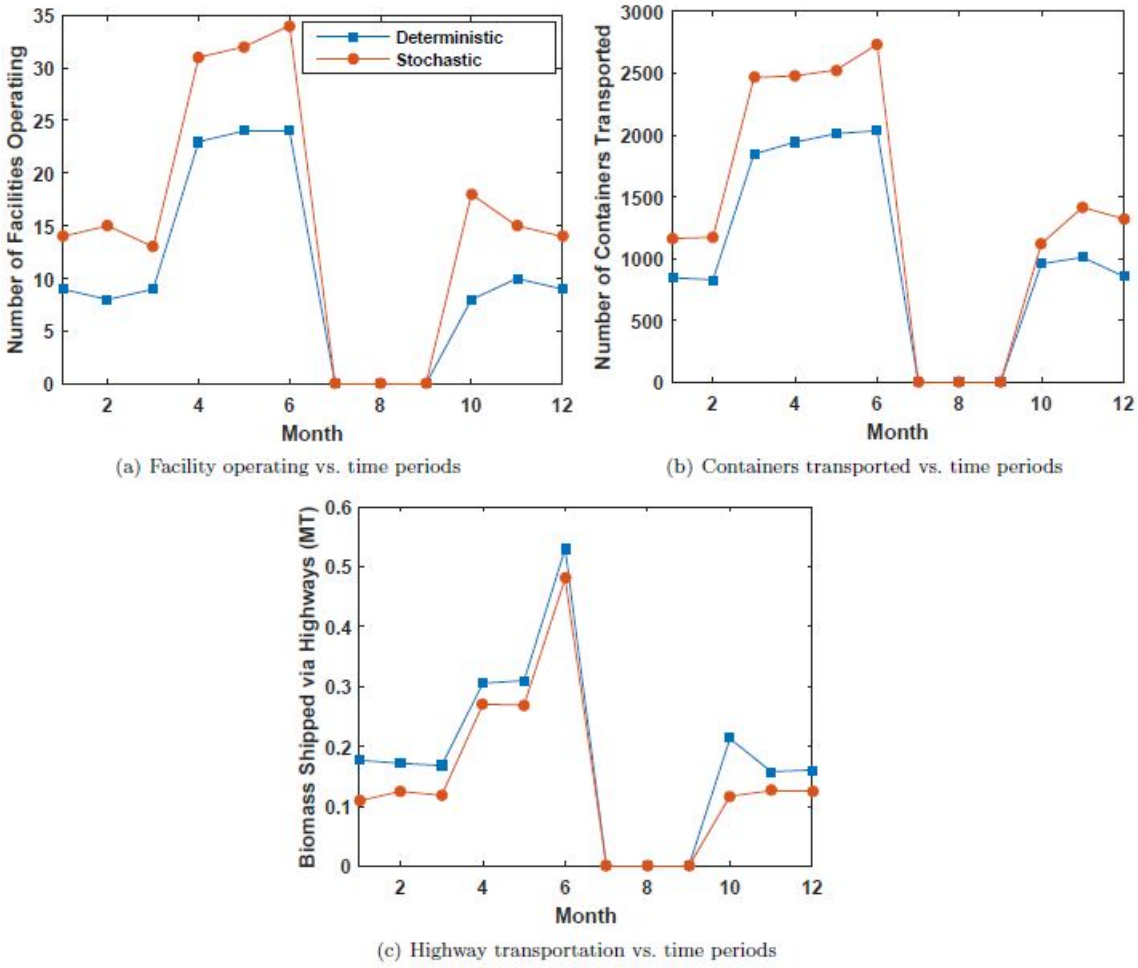
4.4.2.3 Performance evaluation of stochastic vs. deterministic solution

To understand the importance of using stochastic programming approach over deterministic approach this study solves model **[CSM]** using both approaches and compared

their solutions. To solve the deterministic approach, this study uses the average biomass supply $\overline{s}_{bit} = s_{bit\omega}/|\Omega|$ where $s_{bit\omega}$ is obtained from the different supply scenario in stochastic programming approach. Figure 4.9 shows the comparison of key supply chain decisions with and without considering biomass supply uncertainty. It is observed that the stochastic programming approach selects almost 50% more multi-modal facilities $|\mathbf{Y}|$ than the deterministic approach. This results an additional 32.82% increase in container transportation $|\mathbf{Z}|$ between the multi-modal facilities and thus the amount of biomass transported via highways is decreased by 19.19%. In overall, the unit delivery cost of biomass transportation drops from \$45.30/ton (deterministic equivalent solution) to \$41.39/ton (stochastic programming solution) if the biomass seasonality and uncertainty are taken into consideration.

In our experiments, we set t = 1 to represent the month of June. Corn stover is typically harvested from September (t = 4) until November (t = 6) in a given calender year. On the other hard, forest residues are harvested all year, except three months during the winter: December to February (t = 7 - 9). This data supports the statement that September (t = 4) to November (t = 6) are the peak biomass production months in a given calender year whereas neither corn stover nor woody biomass are available during three months of the winter (t = 7 - 9) and can be classified as the off-peak production season for biomass. Therefore, the multi-modal facilities are most likely become congested during the peak biomass production season of a year (t = 4 - 6). We now compare the solutions between deterministic and stochastic approach under high (shown in Figure 4.10) and low biomass (shown in Figure 4.11) availability seasons. It is observed that under both high and low biomass production

seasons stochastic programming solution selects some additional multi-modal facilities to cope against high feedstock seasonality and uncertainty. In summary, we can conclude that feedstock supply uncertainty posses high impact on biomass co-firing supply chain network decisions and use of stochastic solution is beneficial over deterministic solution.



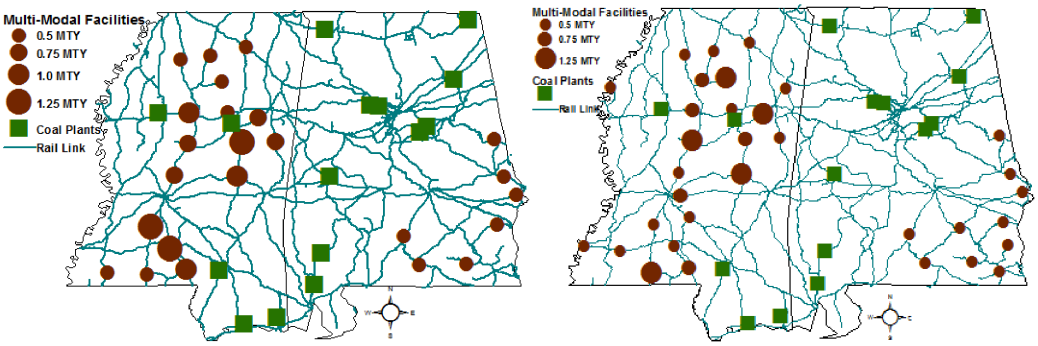
(a) Facility operating vs. time periods

(b) Containers transported vs. time periods

(c) Highway transportation vs. time periods

Figure 4.9

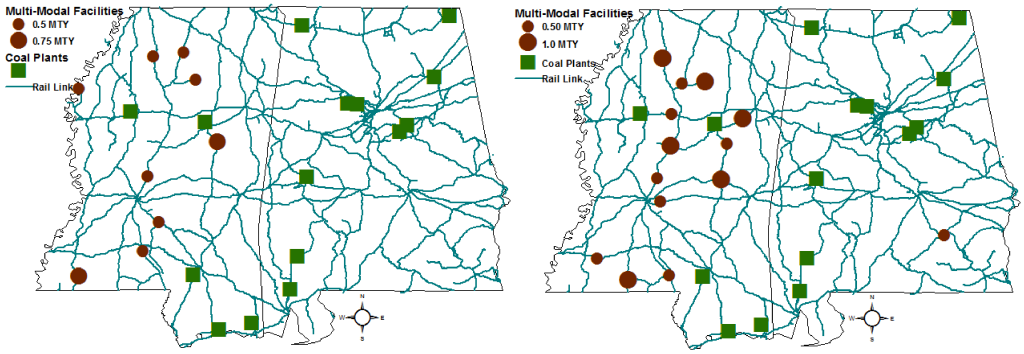
Impact of stochastic vs. deterministic solution on system performance



(a) Deterministic solution (b) Stochastic solution

Figure 4.10

Network representation with and without considering biomass supply uncertainty on peak biomass season (t = 5)



(a) Deterministic solution (b) Stochastic solution

Figure 4.11

Network representation with and without considering biomass supply uncertainty on low biomass season (t = 8)

4.4.3 Analyzing the performance of solution algorithms

We used the algorithms proposed in Section 4.3 to solve model **[LCSM]**. To help the readers follow our approaches, we have used the following notations to represent the algorithms:

- **[CG]**: Constraint Generation Algorithm (CG) algorithm (described in Section 4.3.1)
- **[Hybrid 1]**: Constraint Generation Algorithm (CG) algorithm where the subproblem of the **[CG]** is solved using Sample Average Approximation (SAA) algorithm (described in Section 4.3.2)
- **[Hybrid 2]**: Constraint Generation Algorithm (CG) algorithm where the subproblem of the **[CG]** is solved using Sample Average Approximation (SAA) algorithm and the subproblem of SAA is solved using Progressive Hedging algorithm (PHA) (described in Section 4.3.3)
- **[Hybrid 3]**: Constraint Generation Algorithm (CG) algorithm where the subproblem of the **[CG]** is solved using Sample Average Approximation (SAA) algorithm and the subproblem of SAA is solved using an enhanced Progressive Hedging algorithm (PHA) (enhancement techniques described in Section 4.3.4.1)
- **[Hybrid 4]**: Constraint Generation Algorithm (CG) algorithm where the subproblem of the **[CG]** is solved using Sample Average Approximation (SAA) algorithm and the subproblem of SAA is solved using an enhanced Progressive Hedging algorithm (PHA) (enhancement techniques described in Section 4.3.4.1, 4.3.4.2, and 4.3.4.3)

The algorithms presented above are terminated when at least one of the following condition is met: (a) the optimality gap (i.e., $\epsilon = |UB - LB|/UB$) falls below a threshold value ϵ = 0.01; or (b) the maximum time limit $time^{max}$ = 10,800 (in CPU seconds) is reached; or (c) the maximum number of iteration $iter^{max}$ = 100 is reached. To terminate the Progressive Heading algorithm, we have used some additional stopping criteria which are described at the end of Section 4.3.3. The size of the deterministic equivalent problem of model **[LCSM]** is presented in Table 4.3.

Table 4.3

Problem size of the test instances

Case	$\|\mathcal{I}\|$	$\|\mathcal{J}\|$	$\|\mathcal{K}\|$	$\|\mathcal{L}\|$	$\|\mathcal{B}\|$	$\|\mathcal{T}\|$	Binary Variables	Integer Variables	Continuous Variables	Total Variables	No. of Constraints
1	99	57	15	4	2	4	912	3,420	689,340	693,672	8,329
2	99	57	15	4	2	8	1,824	6,840	1,378,680	1,387,344	16,657
3	99	57	15	4	2	12	2,736	10,260	2,068,020	2,081,016	24,985

Table 4.4.3 presents the results from solving model **[LCSM]** using the algorithms proposed in Section 4.3. We vary the sample size N and then solve model **[LCSM]** with different cases as shown in Table 4.3. Results indicate that, **[CG]** is able to solve only 2 out of 18 problem instances in less than 1% optimality gap within the specified time limit. Introducing SAA to solve the subproblem of **[CG]** (termed as **[Hybrid 1]**) helps to solve 8 out of 18 problem instances in less than 1% optimality gap within the specified time limit. Note that the size of the large scenario set is chosen equal to N' = 500 to evaluate the SAA gap. We then further improve the performance of the **[Hybrid 1]** algorithm by introducing PHA algorithm to solve the subproblems of the SAA algorithm, which we refer as **[Hy-**

brid 2] algorithm. Results further show that **[Hybrid 2]** algorithm is capable of solving 10 out of 18 problem instances in less than 1% optimality gap within the specified time limit. The benefits of using PHA algorithm becomes more evident after applying the enhancement techniques discussed in Section 4.3.4. Results indicate that algorithm **[Hybrid 4]** (employs all the enhancement techniques discussed in Section 4.3.4) saves 30.44% time while dropping the optimality gap by 55.28% over algorithm **[Hybrid 3]** (employs only "*penalty parameter updating scheme*" discussed in Section 4.3.4.1). In summary, **[Hybrid 4]** algorithm seems to offer high quality solutions consistently within the experimental range.

4.5 Conclusion

This paper studies the impact of multi-modal facility congestion on the design and management of a biomass co-firing supply chain network under feedstock supply uncertainty. A two-stage stochastic mixed-integer nonlinear programming model **[CSM]** is developed that not only determines the optimal locations of multi-modal facilities, feedstock storage and processing plans, and shipment routes for biomass transportation but also considers delay costs associated with hub congestions under feedstock supply uncertainty. We then develop a hybrid decomposition algorithm that combines Constraint Generation algorithm with Sample Average Approximation algorithm and an enhanced Progressive Hedging algorithm to solve problem **[CSM]**. The enhanced Progressive Hedging algorithm incorporates several algorithmic improvements such as the dynamic penalty parameter updating technique, the local and global heuristic techniques, and the Rolling horizon algorithm.

Table 4.4

Comparison of different solution approaches

N	Case	[CG] Gap (%)	[CG] CPU (sec)	[CG] Iter	[Hybrid 1] Gap (%)	[Hybrid 1] CPU (sec)	[Hybrid 1] Iter	[Hybrid 2] Gap (%)	[Hybrid 2] CPU (sec)	[Hybrid 2] Iter	[Hybrid 3] Gap (%)	[Hybrid 3] CPU (sec)	[Hybrid 3] Iter	[Hybrid 4] Gap (%)	[Hybrid 4] CPU (sec)	[Hybrid 4] Iter
Normal distribution																
20	1	0.32	6,442	1	0.25	384	1	0.61	**258**	1	0.75	264	2	0.12	308	2
	2	35.46	10,800	1	0.73	1,486	1	0.23	1,284	1	0.11	1,294	1	0.42	**1,239**	2
	3	*out of memory*[a]			*out of memory*			*out of memory*			0.97	9,627	1	0.23	**2,785**	1
30	1	0.63	4,496	1	0.36	1,065	1	0.85	**1,043**	1	0.84	6,963	1	0.46	5,876	1
	2	*out of memory*			*out of memory*			3.15	10,800	1	3.45	10,800	1	**1.21**	9,548	1
	3	*out of memory*			*out of memory*			*out of memory*			3.58	10,800	1	**1.03**	10,800	2
50	1	*out of memory*			0.87	9,458	1	0.59	10,256	1	0.82	4,288	1	0.18	**1,240**	2
	2	*out of memory*			*out of memory*			0.38	10,267	1	0.26	10,568	1	0.40	**3,922**	1
	3	*out of memory*			*out of memory*			*out of memory*			2.12	10,800	1	**1.78**	10,800	1
Average		12.1[b]	7,246	1.0	0.31	3,098	1	0.96	5,651	1	1.43	7,267	1.1	0.65	**5,169**	1.4
Uniform distribution																
20	1	2.12	10,800	3	0.18	4,379	2	0.15	4,561	2	0.34	3,844	2	0.12	**775**	1
	2	25.32	10,800	2	0.97	5,042	1	0.56	5,128	2	0.62	4,743	1	0.20	**1,362**	2
	3	*out of memory*			*out of memory*			21.23	10,800	1	1.28	10,800	1	0.94	**6,458**	1
30	1	9.46	10,800	2	0.42	4,326	2	0.21	4,159	2	0.12	7,827	2	0.13	**1,228**	2
	2	*out of memory*			*out of memory*			0.35	**10,468**	1	2.14	10,800	1	0.88	10,800	1
	3	*out of memory*			*out of memory*			*out of memory*			4.12	10,800	1	**1.64**	10,800	1
50	1	*out of memory*			0.86	6,184	2	0.46	5,248	1	0.48	9,456	1	0.67	**1,276**	1
	2	*out of memory*			*out of memory*			*out of memory*			1.89	10,800	1	**1.20**	10,800	1
	3	*out of memory*			*out of memory*			*out of memory*			6.24	10,800	1	**2.42**	10,800	1
Average		12.30[b]	10,800	2.3	0.61	4,983	1.8	3.83	6,727	1.5	1.91	8,874	1.2	0.84	**6,033**	1.2

[a]CPLEX runs out of memory

[b]Instances with (a) did not contribute to average calculations

Computational results indicate that the hybrid decomposition algorithm (**[Hybrid 4]**) is capable of producing high quality solutions in a reasonable amount of time to solve realistic large-size problem instances for model **[CSM]**.

We use the states of Mississippi and Alabama as a testing ground to evaluate the performance of the modeling results for our study. Our computational experiments reveal some insightful results about the impact of facility congestion on a biomass co-fired supply chain network under feedstock supply uncertainty. We observe that the congestion cost plays a significant role in selecting the size and location of the multi-modal facilities in a biomass co-fired supply chain network. Moreover, we observe that the unit biomass transportation cost drops from $45.30/ton (deterministic equivalent solution) to $41.39/ton (stochastic programming solution) if the biomass seasonality and uncertainty are taken into consideration. The sensitivity analyses further reveal the impact of biomass supply changes on biomass co-fired supply chain network performance.

This work can be extended in several directions. Our study assumes that the congestion occurs only in multi-modal facilities. However, in reality congestion may occur in both highways and in multi-modal facilities. Moreover, our study assumes that the multi-modal facilities are reliable and will never fail. It will be interesting to see how biomass supply chain mitigates congestion under facility disruptions. These issues will be addressed in future studies.

REFERENCES

[1] An H., Wilhelm W.E., Searcy S.W., "A mathematical model to design a lignocellulosic bio-fuel supply chain system with a case study based on a region in central Texas.," *Bioresource Technology*, vol. 102, 2011, pp. 7860–7870.

[2] Association of American Railroads., "National Rail Infrastructure Capacity and Investment Study.," Available from: `http://ops.fhwa.dot.gov/freight/freight_analysis/freight_story/congestion.htm`, 2005.

[3] Awudu A., Zhang J., "Stochastic production planning for a biofuel supply chain under demand and price uncertainties.," *Applied Energy*, vol. 103, 2013, pp. 189–196.

[4] Y. Bai, T. Hwang, S. Kang, and Y. Ouyang, "Biofuel refinery location and supply chain planning under traffic congestion," *Transportation Research Part B: Methodological*, vol. 45, no. 1, 2011, pp. 162 – 175.

[5] Bai Y., Hwang T., Kang S., Ouyang Y., "Biofuel refinery location and supply chain planning under traffic congestion.," *Transportation Research Part B: Methodological*, vol. 45, no. 1, 2011, pp. 162–175.

[6] Bai Y., Li X., Peng F., Wang X., Ouyang Y., "Effects of disruption risks on biorefinery location design.," *Energies*, vol. 8, no. 2, 2015, pp. 1468–1486.

[7] Balasubramanian J., Grossmann I., "Approximation to multistage stochastic optimization in multiperiod batch plant scheduling under demand uncertainty.," *Industrial & Engineering Chemistry Research*, vol. 43, no. 14, 2004, pp. 3695–3713.

[8] Bana e Costa C.A., Oliveira C.S., Vieira V., "Prioritization of bridges and tunnels in earthquake risk mitigation using multicriteria decision analysis: application to Lisbon.," *Omega*, vol. 36, no. 3, 2008, pp. 442–450.

[9] Basic Open-source Mixed INteger programming (BONMIN), ," Available from: `http://projects.coin-or.org/Bonmin`, 2015.

[10] Bioenergy Knowledge Discovery Framework (KDF), "Available from: `https://bioenergykdf.net/taxonomy/term/1036`,", 2013.

[11] Brower M., "Woody biomass Economics.," Available from: http://www.mosaicllc.com/documents/34.pdf, 2010.

[12] Camargo R.S., Miranda Jr. G., Ferreira R., Luna H.P., "Multiple allocation hub and spoke network design under hub congestion.," *Computers & Operations Research*, vol. 36, no. 12, 2009, pp. 3097–3106.

[13] Canel C., Khumawala B.M., Law J., Loh A., "An algorithm for the capacitated, multi-commodity multi-period facility location problem.," *Computers & Operations Research*, vol. 28, no. 5, 2001, pp. 411–427.

[14] Carpentier P.L., Gendreau M., Bastin F., "Long-term management of a hydroelectric multireservoir system under uncertainty using the progressive hedging algorithm," *Water Resources Research*, vol. 49, no. 5, 2013, pp. 2812–2827.

[15] Carrizosa E., Nickel S., "Robust facility location," *Mathematical Methods in Operations Research*, vol. 58, 2003, pp. 331–349.

[16] Chang M., Tseng Y., Chen J., "A scenario planning approach for the flood emergency logistics preparation problem under uncertainty.," *Transportation Research Part E: Logistics and Transportation Review*, vol. 43, no. 6, 2007, pp. 737–754.

[17] Chen C.-W., Fan Y., "Bioethanol supply chain system planning under supply and demand uncertainities.," *Transportation Research Part E*, vol. 48, 2012, pp. 150–164.

[18] Clark S., Watling D., "Modeling network travel time reliability under stochastic demand.," *Transportation Research Part B*, vol. 39, no. 2, 2005, pp. 119–140.

[19] Crainic T.G., Fu X., Gendreau M., Rei W., Wallace S.W., "Progressive hedging-based metaheuristics for stochastic network design," *Networks*, vol. 58, 2011, pp. 114–124.

[20] Credeur M.J., "FedEx in Paris disrupted as workers walk out," Available from: `http://www.bloomberg.com/news/2011-05-13.html`, 2011.

[21] N. Cressie, *Statistics for spatial data*, vol. 4, Wiley Online Library, 1992.

[22] Cui T., Ouyang Y., Shen Z.M., "Reliable facility location under the risk of disruptions.," *Operations Research*, vol. 58, no. 4, 2010, pp. 998–1011.

[23] Cundiff J.S., Dias N., Sherali H.D., "A linear programming approach for designing a herbaceous biomass delivery system.," *Bioresource Technology*, vol. 59, 1997, pp. 47–55.

[24] Daskin M.S., "Application of an expected covering model to emergency medical service system design.," *Decision Sciences*, vol. 13, no. 3, 1982, pp. 416–439.

[25] Daskin M.S., "A maximum expected covering location model: formulation, properties and heuristic solution.," *Transportation Science*, vol. 17, no. 1, 1983, pp. 48–70.

[26] Demirbas A., "Sustainable cofiring of biomass with coal.," *Energy Conservation and Management*, vol. 44, no. 9, 2003, pp. 1465–1479.

[27] Demirbas A., "Biofuels from Agricultural Biomass.," *Energy Sources, Part A*, vol. 31, 2009, pp. 1573–1582.

[28] Drezner Z., "Heuristic solution methods for two location problems with unreliable facilities.," *Journal of the Operational Research Society*, vol. 38, no. 6, 1987, pp. 509–514.

[29] Eksioglu S.D., Acharya A., Leightley L.E., Arora S., "Analyzind the design and management of biomass-to-biorefinery supply chain.," *Computers & Industrial Engineering*, vol. 57, 2009, p. 13421352.

[30] Eksioglu S.D., Li S., Zhang S., Sokhansanj S., Petrolia D., "Analyzing impact of intermodal facilities on design and management of bio-fuel supply chain.," *Transportation Research Record*, vol. 2191, 2010, pp. 144–151.

[31] Eksioglu S.D.,Karimi H., and Eksioglu B., "Optimization Models to Integrate Production and Transportation Planning for Biomass Co-Firing in Coal-Fired Power Plants.," *IIE Transactions*, vol. 0, no. ja, 2016, pp. 0–0.

[32] S. Elhedhli and F. X. Hu, "Hub-and-spoke Network Design with Congestion," *Comput. Oper. Res.*, vol. 32, no. 6, June 2005, pp. 1615–1632.

[33] Elhedhli S., Wu, H., "A Lagrangean Heuristic for Hub-and-Spoke System Design with Capacity Selection and Congestion.," *INFORMS Journal on Computing*, vol. 22, no. 2, 2010, pp. 282–296.

[34] Farahani R., Drezner Z., Asgari N., "Single facility location and relocation problem with time dependent weights and discrete planning horizon.," *Annals of Operations Research*, vol. 167, no. 1, 2009, pp. 353–368.

[35] Fischetti M., Lodi A., "Local branching.," *Mathematical Programming*, vol. 98, 2003, pp. 23–47.

[36] Galindo G., Batta R., "Prepositioning of supplies in preparation for a hurricane under potential destruction of prepositioned supplies," *Socio-Economic Planning Sciences*, vol. 47, 2013, pp. 20–37.

[37] GAMS: CONOPT Solver Manual, ," Available from: `http://gams.com/dd/docs/solvers/conopt/index.html`, 2015.

[38] Gebreslassie B.H., Yao Y., You F., "Design under uncertainty of hydrocarbon biorefinery supply chains: Multiobjective stochastic programming models, decomposition algorithm, and a comparison between CVaR and downside risk.," *AIChE Journal*, vol. 58, no. 7, 2012, p. 21552179.

[39] General Algebraic Modeling System (GAMS), ," Available from: `http://www.gams.com/`, 2013.

[40] Geoffrion A.M., "Generalized Benders decomposition.," *Journal of Optimization Theory and Applications*, vol. 10, 1972, pp. 237–260.

[41] Giarola S., Zamboni A., Bezzo F., "Spatially explicit multi-objective optimization for designing and planning of hybrid fast and second generation biorefineries.," *Computers and Chemical Engineering*, vol. 35, 2011, pp. 1782–1797.

[42] Gonzales D., Searcy E.M., Eksioglu S.D., "Cost analysis for high-volume and long-haul transportation of densified biomass feedstock.," *Transportation Research Part A*, vol. 49, 2013, pp. 48–61.

[43] Grove G.P., OKelly M.E., "Hub networks and simulated schedule delay.," *Papers of the Regional Science Association*, vol. 59, 1986, pp. 103–119.

[44] Gul S., Denton B.T., Fowler J., "A Multi-Stage Stochastic Integer Programming Model for Surgery Planning," *INFORMS Journal on Computing*, vol. In-press, 2015.

[45] L. Hajibabai and Y. Ouyang, "Integrated Planning of Supply Chain Networks and Multimodal Transportation Infrastructure Expansion: Model Development and Application to the Biofuel Industry," *Computer-Aided Civil and Infrastructure Engineering*, vol. 28, no. 4, 2013, pp. 247–259.

[46] Haugen K.K., and Lokkentangen A., "Progressive hedging as a metaheuristic applied to stochastic lot-sizing.," *European Journal of Operational Research*, vol. 132, 2001, pp. 116–122.

[47] Helgason T., Wallace S.W., "Approximate scenario solutions in the progressive hedging algorithm," *Annals of Operations Research*, vol. 31, 1991, pp. 425–444.

[48] Hess J.R., Wright T.C., Kenney L.K., Searcy M.E., "Uniform-format solid feedstock supply system: A commodity-scale design to produce an infrastructure-compatible bulk solid from lignocellulosic biomass.," INL/EXT-09-15423. Available from: `http://www.inl.gov/technicalpublications/Documents/4408280.pdf`, 2009.

[49] Huang Y., Chen C.W., Fan Y., "Multistage optimization of the supply chains of bio-fuels.," *Transportation Research Part E*, vol. 46, no. 6, 2010, pp. 820–830.

[50] Huang Y., Fan Y., Chen C.-W, "An integrated bio-fuel supply chain against feedstock seasonality and uncertainty.," *Transportation Science*, vol. 48, no. 4, 2014, pp. 540–554.

[51] Hubbard R., "BNSF Railway to put $6 billion toward relieving congestion.," Available from: http://www.omaha.com/money/bnsf-railway-to-put-billion-toward-relieving-congestion/article_f00f7b1d-353e-58e3-abd5-3967d8f2c474.html, 2014.

[52] Hurricanes: Science and Society, "Rainfall and Inland Flooding," Available from: http://www.hurricanescience.org/society/impacts/rainfallandinlandflooding/, 2013.

[53] Hvattum L.M., Lokketangen A., "Using scenario trees and progressive hedging for stochastic inventory routing problems," *Journal of Heuristics*, vol. 15, 2009, pp. 527–557.

[54] A. I. and Z. J., "Uncertainty and sustainability concepts in biofuel supply chain management: A review.," *Renewable and Sustainable Energy Reviews*, vol. 16, 2012, pp. 1359–1368.

[55] Jena S.D., Cordeau J.F., Gendron B., "Solving a Dynamic Facility Location Problem with Partial Closing and Reopening.," *Computers & Operations Research*, vol. In-press., 2015.

[56] Kara B.Y., Tansel B.C., "The latest arrival hub location problem.," *Management Science*, vol. 47, no. 10, 2001, pp. 1408–1420.

[57] Karimi H., Eksioglu S.D., and Hu M., "A model for analyzing the impact of production tax credit on renewable electricity production.," Proceedings of the Industrial & Systems Engineering Research Conference (ISERC), May 31-June 3, Montreal, Canada, 2014.

[58] Keyvanshokooh E., Ryan S.M., Kabir E., "Hybrid robust and stochastic optimization for closed-loop supply chain network design using accelerated Benders decomposition.," *European Journal of Operational Research*, vol. 249, no. 1, 2016, pp. 76–92.

[59] Kim J., Realff M.J., Lee J.H., "Optimal design and global sensitivity analysis of biomass supply chain networks for bio-fuels under uncertainty.," *Computers & Chemical Engineering*, vol. 35, 2011, pp. 1738–1751.

[60] Kleywegt A.J., Shapiro A., Homem-De-Mello T., "The sample average approximation method for stochastic discrete optimization.," *SIAM Journal of Optimization*, vol. 12, 2001, pp. 479–502.

[61] Klose A., Drexl A., "Facility location models for distribution system design," *European Journal of Operational Research*, vol. 162, no. 1, 2005, pp. 4–29.

[62] Kostina A.M., Guillen-Gosalbeza G., Meleb F.D., Bagajewiczc M.J., Jimeneza L., "A novel rolling horizon strategy for the strategic planning of supply chains. Application to the sugar cane industry of Argentina.," *Computers & Chemical Engineering*, vol. 35, 2011, pp. 2540–2563.

[63] Lee D.H., Dong M., "Dynamic network design for reverse logistics operations under uncertainty.," *Transportation Research Part E*, vol. 45, no. 1, 2009, pp. 61–71.

[64] Li Q., Zeng B., Savachkin A., "Reliable facility location design under disruptions.," *Computers & Operations Research*, vol. 40, 2013, pp. 901–909.

[65] Li X., Ouyang Y., "A continuous approximation approach to reliable facility location design under correlated probabilistic disruptions.," *Transportation Research Part B*, vol. 44, 2010, pp. 535–548.

[66] Li X., Peng F., Bai Y., Ouyang Y., "Effects of disruption risks on biorefinery location design: Discrete and continuous models.," *proceeding of the* 90^{th} *TRB Annual Meeting, Washington D.C.*, January 2011.

[67] Lo H.K., Luo X.W., Siu B.W.Y., "Degradable transport network: travel time budget of travelers with heterogeneous risk aversion.," *Transportation Research Part B*, vol. 40, no. 9, 2006, pp. 792–806.

[68] Lo H.K., Tung Y.K., "Network with degradable links: capacity analysis and design.," *Transportation Research Part B*, vol. 37, no. 4, 2003, pp. 345–363.

[69] Magnanti T.L., Wong R.T., "Acclerating Benders Decomposition: Algorithmic enhancement and model selection criteria.," *Operations Research*, vol. 29, 1981, pp. 464–484.

[70] Mahmudi H., Flynn P., "Rail vs. truck transport of biomass.," *Applied Biochemistry and Biotechnology*, vol. 129, no. 1, 2006, pp. 88–103.

[71] Mak W.K., Morton D.P., Wood R.K., "Monte Carlo bounding techniques for determining solution quality in stochastic programs.," *Operations Research Letters*, vol. 24, 1999, pp. 47–56.

[72] Marianov V., Serra D., "Location models for airline hubs behaving as M/D/c queues.," *Computers and Operations Research*, vol. 30, 2003, pp. 983–1003.

[73] Marufuzzaman M., Eksioglu S.D., "Designing a Reliable and Dynamic Multi-Modal Transportation Network for Biofuel Supply Chain.," *Transportation Science*, 2016.

[74] Marufuzzaman M., Eksioglu S.D., Hernandez R., "Environmentally friendly supply chain planning and design for biodiesel production via wastewater sludge.," *Transportation Science*, vol. 48, no. 4, 2014, pp. 555–574.

[75] Marufuzzaman M., Eksioglu S.D., Huang Y., "Two-stage stochastic programming supply chain model for biodiesel production via wastewater treatment.," *Computers & Operations Research*, vol. 49, 2014, pp. 1–17.

[76] Marufuzzaman M., Eksioglu S.D., Li X., Wang J., "Analyzing the impact of intermodal-related risk to the design and management of bio-fuel supply chain.," *Transportation Research Part E*, vol. 69, 2014, pp. 122–145.

[77] Melkote S., Daskin M.S., "Capacitated facility location-network design problems," *European Journal of Operational Research*, vol. 129, no. 3, 2001, pp. 481–495.

[78] Melo M.T., Nickel S., Saldanha-da-Gama F., "Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning," *Computers & Operations Research*, vol. 33, 2005, pp. 181–208.

[79] Melo M.T., Nickel S., Saldanha-da-Gama F., "Facility location and supply chain management A review," *European Journal of Operational Research*, vol. 196, 2009, pp. 401–412.

[80] Memisoglu G., Uster H., "Integrated bio-energy supply chain network planning problem.," *Transportation Science*, vol. In Press, 2015.

[81] Meyer A.D., Cattrysse D., Orshoven J.V., "A generic mathematical model to optimise strategic and tactical decisions in biomass-based supply chains (OPTIMASS)," *European Journal of Operational Research*, vol. 245, no. 1, 2015, pp. 247–264.

[82] Miranda Jr. G.,, de Camargo R.S., Pinto L.R., Conceicao S.V., Ferreira R.P.M, "Hub location under hub congestion and demand uncertainty: The Brazilian case study.," *Pesquisa Operacional*, vol. 31, no. 2, 2011, p. 319349.

[83] Moghtaderi-Zadeh M., Der Kiureghian A., "Reliability upgrading of lifeline networks for post-earthquake serviceability.," *Earthquake Engineering and Structural Dynamics*, vol. 11, no. 4, 1983, pp. 557–566.

[84] Mulvey J.M., Vladimirou H., "Applying the progressive hedging algorithm to stochastic generalized networks," *Annals of Operations Research*, vol. 31, 1991, pp. 399–424.

[85] National Renewable Energy Laboratory, "Biomass Cofiring A renewable alternative for utilities," Available from: `http://www.nrel.gov/docs/fy00osti/28009.pdf`, 2000.

[86] Norkin V.I., Ermoliev Y.M., Ruszczynski A., "On optimal allocation of indivisibles under uncertainty.," *Operations Research*, vol. 46, 1998, pp. 381–395.

[87] Norkin V.I., Pflug G.C., Ruszczynski A., "A branch and bound method for stochastic global optimization.," *Mathematical Programming*, vol. 83, no. 3, 1998, pp. 425–450.

[88] Parker N., Tittmann P., Hart Q., Lay M., Cunningham J., Jenkins B., "Strategic assessment of bioenergy development in the west: Spatial analysis and supply curve development.," *Western Governors' Association*, 2008.

[89] Peeta S., Salman F.S., Gunnec D., Viswanath K., "Pre-disaster investment decisions for strengthening a highway network.," *Computers & Operations Research*, vol. 37, 2010, pp. 1708–1719.

[90] Persson T., Garcia A., Paz J., Jones J., Hoogenboom G., "Maize ethanol feedstock production and net energy value as affected by climate variability and crop management practices.," *Agricultural Systems*, vol. 100, no. 1-3, 2009, pp. 11–21.

[91] Polson J., "BP oil still ashore one year after end of Gulf spill," Available from: `http://www.bloomberg.com/news/2011-07-15.html`, 2011.

[92] Poudel S., Marufuzzaman M., Bian L., "Designing a reliable biofuel supply chain network considering link failure probabilities.," *Computers & Industrial Engineering*, vol. 91, 2016, pp. 85–99.

[93] Poudel S.R, Marufuzzaman M., Bian L., "A hybrid decomposition algorithm for designing a multi-modal transportation network under biomass supply uncertainty.," *Transportation Research Part E: Logistics and Transportation Review*, vol. 94, 2016, pp. 1–25.

[94] Poudel S.R, Marufuzzaman M., Eksioglu S.D., AmirSadeghi M., French T., "Supply chain network model for biodiesel production via wastewaters from paper and pulp companies.," *Handbook of Bioenergy*, 2015, pp. 163–190.

[95] Rajgopal J., Wang Z., Schaefer A.J., Prokopyev O.A., "Integrated design and operation of remnant inventory supply chains under uncertainty.," *European Journal of Operational Research*, vol. 214, no. 2, 2011, pp. 358–364.

[96] Rawls C.G., Turnquist M.A., "Pre-positioning of emergency suppliers for disaster response," *Transportation Research Part B*, vol. 44, 2010, pp. 521–534.

[97] Rawls C.G., Turnquist M.A., "Pre-positioning planning for emergency response with service quality constraints," *OR Spectrum*, vol. 33, 2011, pp. 481–498.

[98] Rawls C.G., Turnquist M.A., "Pre-positioning and dynamic delivery planning for short-term response following a natural disaster," *Socio-Economic Planning Sciences*, vol. 46, 2012, pp. 46–54.

[99] Rockafellar R.T., Wets R.J.-B., "Scenarios and policy aggregation in optimization under uncertainty," *Mathematics of operations research*, vol. 16, 1991, pp. 119–147.

[100] Roni M.S, Eksioglu S.D., Searcy E., Jha K., "A supply chain network design model for biomass co-firing in coal-fired power plants.," *Transportation Research Part E*, vol. 61, 2014, pp. 115–134.

[101] Santos M.L.L., da Silva E.L., Finardi E.C., Goncalves R.E.C., "Practical aspects in solving the medium-term operation planning problem of hydrothermal power systems by using the progressive hedging method," *International Journal of Electrical Power and Energy Systems*, vol. 31, 2009, pp. 546–552.

[102] Santoso T., Ahmed S., Goetschalckx M., Shapiro A., "A stochastic programming approach for supply chain network design under uncertainty.," *European Journal of Operational Research*, vol. 167, 2005, pp. 96–115.

[103] Schutz P., Tomasgard A., Ahmed S., "Supply chain design under uncertainty using sample average approximation and dual decomposition.," *European Journal of Operational Research*, vol. 199, 2009, pp. 409–419.

[104] Shen J.M., Zhan R.L., Zhang J., "The reliable facility location problem: Formulations, heuristics, and approximation algorithms.," *INFORMS Journal on Computing*, vol. 23, no. 3, 2011, pp. 470–482.

[105] Shiode S., Drezner Z., "A competitive facility location problem on a tree network with stochastic weights," *European Journal of Operational Research*, vol. 149, 2003, pp. 47–52.

[106] Snyder L.V., "Facility Location Under Uncertainty: A Review," *IIE Transactions*, vol. 38, no. 7, 2006, pp. 537–554.

[107] Snyder L.V., Daskin M.S., "Reliability models for facility location: The expected failure cost case.," *Transportation Science*, vol. 39, no. 3, 2005, pp. 400–416.

[108] Sohn J., "Evaluating the significance of highway network links under the flood damage: an accessibility approach.," *Transportation Research Part A*, vol. 40, no. 6, 2006, pp. 491–506.

[109] SteadieSeifi M., Dellaert N.P., Nuijten W., Woensel T.V., Raoufi R., "Multimodal freight transportation planning: a literature review.," *European Journal of Operational Research*, vol. 233, 2014, pp. 1–15.

[110] The National Energy Technology Laboratory, "Coal-Fired Power Plants in the United States.," Available from: http://www.netl.doe.gov/energyanalyses/hold/technology.html, 2005.

[111] The U.S. Department of Agriculture, "World Agricultural Outlook Board: World Agricultural Supply and Demand Estimates and supporting materials.," Available from: http://www.usda.gov/oce/commodity/wasde/, 2015.

[112] Torres-Soto J.E., Uster H., "Dynamic-demand capacitated facility location problems with and without relocation.," *International Journal of Production Research*, vol. 49, no. 13, 2011, pp. 3979–4005.

[113] United States Energy Information Administration, "State Energy Data System (SEDS): 2012 (Updates)," Available from: http://www.eia.gov/state/seds/seds-data-fuel.cfm?sid=US, 2014.

[114] U.S. Energy Information Administration., "International Energy Statistics.," Available from: http://www.eia.gov/cfapps/ipdbproject/iedindex3.cfm?tid=79\&pid=80\&aid=1\&cid=regions\&syid=2000\&eyid=2011\&unit=TBPD, 2013.

[115] U.S. EPA., "United States Environmental Protection Agency: Renewable Fuel Standard 467 Implementation.," Available from: http://www.epa.gov/OTAQ/renewablefuels/index.htm, 2013.

[116] Verweij B., Ahmed S., Kleywegt A.J., Nemhauser G., Shapiro A., "The sample average approximation method applied to stochastic routing problems: A computational study.," *Computational Optimization and Applications*, vol. 24, 2003, pp. 289–333.

[117] N. Vidyarthi and S. Jayaswal, "Efficient solution of a class of locationallocation problems with stochastic demand and congestion," *Computers and Operations Research*, vol. 48, 2014, pp. 20 – 30.

[118] Vidyarthi N., Jayaswal S., "Efficient solution of a class of locationallocation problems with stochastic demand and congestion.," *Computers and Operations Research*, vol. 48, 2014, pp. 20–30.

[119] Wallace S.W., Helgason T., "Structural properties of the progressive hedging algorithm," *Annals of Operations Research*, vol. 31, 1991, pp. 445–456.

[120] Walther G., Schakta B., Spengler T.S., "Design of regional production networks for second generation synthetic bio-fuel A case study in Northern Germany.," *European Journal of Operational Research*, vol. 218, 2012, pp. 280–292.

[121] Walther G., Schatka A., Spengler T.S., "Design of regional production networks for second generation synthetic bio-fuel A case study in Northern Germany," *European Journal of Operational Research*, vol. 218, no. 1, 2012, pp. 280–292.

[122] Wang X., Ouyang Y., "A continuous approximation approach to competitive facility location design under facility disruption risks.," *Transportation Research Part B*, vol. 50, 2013, pp. 90–103.

[123] Watson J.P., Woodruff D.L., "Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems," *Computational Management Science*, vol. 8, 2011, pp. 355–370.

[124] Weather Underground, Inc., "A Detailed View of the Storm Surge: Comparing Katrina to Camille.," Available from: `http://www.wunderground.com/hurricane/surge_details.asp?`, 2014.

[125] Weber A., *Uber den standort der industrian,translated as (in 1929): Alferd weber's theory of the location of industries*, University of Chicago press, 1909.

[126] Williams J. L., "Information theoretic sensor management.," Available from: `http://dspace.mit.edu/handle/1721.1/38534`, 2007.

[127] Xie F., Huang Y., Eksioglu S. , "Integrating multimodal transport into cellulosic biofuel supply chain design under feedstock seasonality with a case study based on California.," *Bioresource Technology*, vol. 152, 2014, pp. 15–23.

[128] Xie F., Huang Y., Eksioglu, S.D., "Integrating multimodal transport into cellulosic bio-fuel supply chain design under feedstock seasonality with a case study based on California.," *Bioresource Technology*, vol. 152, 2014, pp. 15–23.

[129] Xie W., Ouyang Y., "Dynamic planning of facility locations with benefits from multitype facility colocation.," *Computer-Aided Civil & Infrastructure Engineering*, vol. 28, no. 9, 2013, pp. 666–678.

[130] You F., Tao L., Graziano D. J., Snyder S. W., "Optimal design of sustainable cellulosic bio-fuel supply chains: Multiobjective optimization coupled with life cycle assessment and input-output analysis.," *AIChE Journal*, vol. 58, no. 4, 2012, p. 11571180.

[131] Zamboni A., Bezzo F., Shah N., "Spatially explicit static model for the strategic design of future bioethanol production systems. 2. Multi-objective environmental optimization.," *Energy & Fuels*, vol. 23, no. 10, 2009, p. 51345143.

[132] Zamboni A., Shah N., Bezzo F., "Spatially explicit static model for the strategic design of future bioethanol production systems.1. Cost Minimization.," *Energy & Fuels*, vol. 23, no. 10, 2009, pp. 5121–5133.