Stochastic Optimization for Renewable Energy System Design and Operation

Berkay Eren Gulcan
Clemson University, berkaygulcan@gmail.com

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Recommended Citation
https://tigerprints.clemson.edu/all_dissertations/2864

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
STOCHASTIC OPTIMIZATION FOR RENEWABLE ENERGY SYSTEM DESIGN AND OPERATION

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Industrial Engineering

by
Berkay Eren Gulcan
August 2021

Accepted by:
Dr. Yongjia Song, Committee Chair
Dr. Sandra D. Eksioglu
Dr. Amin Khademi
Dr. Jeffrey P. Kharoufeh
Abstract

In this dissertation, we present novel stochastic optimization models and solution methods for the optimization of renewable energy systems. Specifically, the focus is on bioenergy supply chains and operations in the presence of uncertainty. Overall, our applications highlight the breadth in which stochastic programming models can be applied in the energy sector. We first study the impact of biomass blending practice on reducing the cost of producing biofuels. In detail, we consider a biomass supply chain problem where the goal is to identify purchase quantities of different biomass feedstocks from possible suppliers in order to meet the quality requirements of the biomass conversion process. The model incorporates the quality requirements using chance constraints that take into account the stochastic nature of biomass quality. There are two problem settings in this study, a centralized and a decentralized supply chain. Proposed is a mixed-integer linear program that models the blending problem in the centralized setting and the bilevel program models the blending problem in the decentralized setting. We use the sample average approximation (SAA) method to approximate the chance constraints and propose solution algorithms to solve this approximation. The case study developed for South Carolina using the Billion Ton Study data provides an environment to conduct numerical experiments. Numerical results demonstrate that the blends identified and the suppliers selected by both models are different, and the cost of the centralized supply chain is 2 to 6% lower. The implications of these results are two-fold. First, these results could lead to improved collaborations in the supply chain. Second, these results provide an estimate of the approximation error from assuming centralized decision-making in the supply chain. In addition, we provide managerial insights based on the identified biomass blends.

We next study the optimization of biorefinery operations under stochastic biomass characteristics and stochastic equipment failure. Variations of physical and chemical characteristics of biomass lead to an uneven flow of biomass in a biorefinery, which reduces equipment utilization and increases operational costs. Uncertainty of biomass supply and high processing costs increase the risk of investing in the U.S.’s cellulosic
biofuel industry. We propose a stochastic programming model to streamline processes within a biorefinery. A chance constraint models the system’s reliability requirement that the reactor is operating at a high utilization rate given uncertain biomass moisture content, particle size distribution, and equipment failure. The model identifies operating conditions of equipment and inventory levels to maintain a continuous biomass flow to the reactor. The Sample Average Approximation method approximates the chance constraint, and a bisection search-based heuristic solves this approximation. A case study is developed using real data collected at Idaho National Laboratory’s biomass processing facility. An extensive computational analysis indicates that sequencing of biomass bales based on moisture level, increasing storage capacity, and managing particle size distribution increases utilization of the reactor and reduces operational costs.

Finally, we extend the previous work on optimizing biorefinery operations to integrate the sequential information flow from moisture sensors into the decision-making process. Integrating the sensory data into the operational decisions in biomass processing will increase its responsiveness to the changing biomass conditions. We propose a multi-stage stochastic programming model that minimizes the expected operational costs by identifying the initial inventory level and creating an operational decision policy for equipment speed settings. These policies take the sensory data and the current biomass inventory level as inputs to dynamically adjust inventory levels and equipment settings according to changes in the biomass’ characteristics. We ensure that a prescribed utilization target of the reactor is consistently achieved by penalizing the violation of the target reactor feeding rate. A case study is developed using data collected at Idaho National Laboratory’s biomass processing facility. We show the value of multi-stage stochastic programming from an extensive computational experiment. Our sensitivity analysis indicates that, by updating the infeed rate of the system, the processing speed of equipment and bale sequencing based on moisture level of biomass improves the processing rate of the reactor and reduces operating costs.
Dedication

To my mother, Füsun Gülcan, my dad, Serdar Gülcan, and my brother Baran Gülcan who support me at any means and selflessly endure not being side by side. I feel your love and support regardless of distance.

To my childhood friends, my best men, Berkay, Berke, Tan, and Sarp, who trust me more than I do myself, and always are on my side. They make me realize the value of true friendship/brotherhood.

To my loving fiancé, Alexandra, who has been the source of my endurance and power to keep going.
Acknowledgments

First and foremost, I would like to express my greatest thanks to my Ph.D. advisor Dr. Yongjia Song, who has done everything he could to help me in every aspect since we started working together. Apart from being an excellent scholar, supervisor, and mentor, he put on the supportive therapist hat for me and helped me a lot to overcome stressful, hard, but learning experiences. Wholeheartedly, I would have never completed my Ph.D. without his support. Without distinction, I would like to express my highest appreciation to Dr. Sandra Duni Eksioglu for her support, guidance, and inspiration throughout my doctoral studies. She inspires me every day with her hardworking and passionate attitude, not only for research and teaching but also with her random act of kindness to give back to the community at all levels. I would like to extend my gratitude to Drs. Amin Khademi and Jeffrey P. Kharoufeh for being a part of my committee and providing invaluable comments for my dissertation. I share my sincere thanks to them for being kind while challenging me to excel with my efforts and research.
# Table of Contents

Title Page ................................................................. i

Abstract ................................................................. ii

Dedication ............................................................... iv

Acknowledgments ......................................................... v

List of Tables .......................................................... viii

List of Figures ........................................................ x

1 Introduction .......................................................... 1
   1.1 Background .................................................... 1
   1.2 Overview of the Dissertation .................................. 6

2 A Stochastic Biomass Blending Problem in Decentralized Supply Chains ............. 8
   2.1 Background .................................................... 8
   2.2 Literature Review ............................................. 10
   2.3 Blending Problem in a Centralized Supply Chain .............. 12
   2.4 Solution Approaches: Centralized Blending Model .......... 14
   2.5 Blending Problem in a Decentralized Supply Chain .......... 19
   2.6 Solution Approaches: Decentralized Blending Model .......... 21
   2.7 Numerical Analysis ............................................ 24

3 Optimization Models for Integrated Biorefinery Operations ............................ 31
   3.1 Background .................................................... 31
   3.2 Literature Review ............................................. 33
   3.3 Problem Description and Formulation ........................ 36
   3.4 Case Study .................................................... 45
   3.5 Numerical Experiments and Sensitivity Analysis .............. 47

4 Adaptive Biomass Processing Operations under Uncertainty ............................ 60
   4.1 Background .................................................... 60
   4.2 Literature Review ............................................. 62
   4.3 Problem Description and Mathematical Formulation .......... 64
   4.4 Solution Methodology: Stochastic Dual Dynamic Programming 72
   4.5 Numerical Experiments and Sensitivity Analysis .............. 74

5 Conclusions and Future Research .................................... 89

Appendices ............................................................. 93
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A  Stochastic Biomass Blending Problem</td>
<td>94</td>
</tr>
<tr>
<td>B  Integrated Biorefinery Operations</td>
<td>107</td>
</tr>
<tr>
<td>Bibliography</td>
<td>120</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Summary of Input Data .................................................. 25
2.2 Biomass Refinery Requirements ........................................ 25
2.3 Summary of the Results of Linear Approximation Model .......... 26
2.4 Evaluating the Single-Level Optimization Model for A Small Size Problem ($\beta = 0.1, \gamma = 0.1$) ........................................ 27
2.5 Comparison of ($\bar{Q}_L$) formulation and the Heuristic Algorithm ($N = 1$) .................................................. 27
2.6 Costs and Biomass Blending Ratios Under Ash and Thermal Content Uncertainties ($\beta = \gamma = 0.3$) ............... 28
2.7 Costs and Biomass Blending Ratios Under Ash and Thermal Content Uncertainties ($\beta = \gamma = 0.2$) ............... 28
2.8 Costs and Biomass Blending Ratios Under Ash and Thermal Content Uncertainties ($\beta = \gamma = 0.3$) ............... 29
2.9 Costs and Biomass Blending Ratios Under Ash and Thermal Content Uncertainties ($\beta = \gamma = 0.2$) ............... 29
3.1 Mathematical Notation .................................................. 39
3.2 Distribution Parameters .................................................. 47
3.3 System’s Performance vs. Reliability ................................ 49
3.4 Summary of Results from MIP Formulation and Algorithm 1 .... 49
3.5 Base-Case Problem: Reactor Utilization & Costs ............... 50
3.6 Base-Case Problem: Inventory Level & Equipment Setting .... 50
3.7 Bale Sequencing: Results of the Sensitivity Analysis ........... 51
3.8 System Performance under Short-duration Equipment Stoppages 53
3.9 System Performance under Long-duration Equipment Failures 53
3.10 Base-case Problem: Increased Storage Capacity ............... 54
3.11 Short-Duration Failures: Increased Storage Capacity ........... 55
3.12 Long-Duration Failures: Increased Storage Capacity ........... 56
3.13 Particle Size: Primary Grinder ...................................... 57
3.14 Particle Uniformity: Primary Grinder ................................ 58
3.15 Particle Size: Secondary Grinder .................................... 58
4.1 In-sample stability test - # of scenarios used (per stage) in the multi-stage SP model ............................................ 76
4.2 In-sample stability test - # of scenarios used in the two-stage SP model .................................................. 77
4.3 Out-of-sample stability test - # of sample paths used for performance evaluation ............................................. 77
4.4 Model performance on the base-case problem .................... 78
4.5 Percentage difference in the inventory level obtained by the two-stage SP model and MV problem in stage 36 to stage 50 for the base-case problem compared to the multi-stage SP model. 79
4.6 Model performances for different length of planning horizons .......... 80
4.7 The performance of the “truncate-and-repeat” strategy .......... 81
4.8 Model performances for different mixing ratios of bales with target rate $2.95 \, dt/hr.$ ............................................. 82
4.9 Model performances for different mixing ratios of bales with target rate $2.72 \, dt/hr.$ ............................................. 82
4.10 Model performances for different mixing ratios of bales with target rate $2.50 \, dt/hr.$ ............................................. 83
4.11 Sensitivity analysis for granularity of decision stages. 

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Solutions Returned by <strong>SAA Algorithm</strong> for ((P) (\beta = \gamma = 0.3; \hat{\beta} = \hat{\gamma} = 0.0))</td>
<td>103</td>
</tr>
<tr>
<td>A2</td>
<td>Solutions Returned by <strong>SAA Algorithm</strong> for ((P) (\beta = \gamma = 0.2; \hat{\beta} = \hat{\gamma} = 0.0))</td>
<td>103</td>
</tr>
<tr>
<td>A3</td>
<td>Solutions Returned by <strong>SAA Algorithm</strong> for ((P) (\beta = \gamma = \hat{\beta} = \hat{\gamma} = 0.3))</td>
<td>104</td>
</tr>
<tr>
<td>A4</td>
<td>Solutions Returned by <strong>SAA Algorithm</strong> for ((P) (\beta = \gamma = \hat{\beta} = \hat{\gamma} = 0.2))</td>
<td>104</td>
</tr>
<tr>
<td>A5</td>
<td>Lower Bounds for ((P) (\beta = \gamma = \hat{\beta} = \hat{\gamma} = 0.3))</td>
<td>104</td>
</tr>
<tr>
<td>A6</td>
<td>Lower Bounds for ((P) (\beta = \gamma = \hat{\beta} = \hat{\gamma} = 0.2))</td>
<td>104</td>
</tr>
<tr>
<td>A7</td>
<td>Lower Bounds for ((P))</td>
<td>104</td>
</tr>
<tr>
<td>A8</td>
<td>Solutions returned by <strong>SAA Algorithm</strong> for (\beta = \gamma = 0.3 (\hat{\beta} = \hat{\gamma} = 0.0))</td>
<td>105</td>
</tr>
<tr>
<td>A9</td>
<td>Solutions returned by <strong>SAA Algorithm</strong> for (\beta = \gamma = 0.2 (\hat{\beta} = \hat{\gamma} = 0.0))</td>
<td>105</td>
</tr>
<tr>
<td>A10</td>
<td>Detailed Mathematical Notation</td>
<td>107</td>
</tr>
<tr>
<td>A11</td>
<td>Energy Consumption and Fixed Costs</td>
<td>117</td>
</tr>
<tr>
<td>A12</td>
<td>Particle Size Distribution Percentiles</td>
<td>117</td>
</tr>
<tr>
<td>A13</td>
<td>Moisture and Dry Matter Changes</td>
<td>117</td>
</tr>
<tr>
<td>A14</td>
<td>Equipment Infeed Rate Limits</td>
<td>117</td>
</tr>
<tr>
<td>A15</td>
<td>Particle size distribution percentiles</td>
<td>118</td>
</tr>
<tr>
<td>A16</td>
<td>Secondary grinder bypass ratios</td>
<td>118</td>
</tr>
<tr>
<td>A17</td>
<td>Regression Analysis Statistics</td>
<td>118</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Supply Chain Design for Thermochemical Conversion .......................... 12
2.2 Supply Curve for Woody Biomass in Pickens County, SC. ..................... 13
2.3 Decentralize vs Centralize - Percent Gap ........................................ 30

3.1 Biomass Feeding Processes of PDU ................................................. 37
3.2 Total Cost vs. Reliability ............................................................. 49
3.3 Inventory Level for Different Sequences ......................................... 51
3.4 Average and Standard Deviation of Biomass in Metering Bin. ............... 57
3.5 Average Biomass Density ............................................................. 59
3.6 Normalized Standard Deviation of Biomass Density in the Metering Bin .. 59

4.1 Scenario tree illustration for multi-stage SP. ................................. 69
4.2 Scenario tree illustration for two-stage approximation to multi-stage SP. ..... 70
4.3 Target violation penalties for base-case problem. ............................. 78
4.4 Percentage difference in the inventory level for the two-stage SP and the MV problem for the base-case problem compared with the multi-stage SP model. ..................... 79
4.5 Sensitivity analysis for initial inventory level. ................................. 86
4.6 Sensitivity analysis for the length of the bale sequencing pattern. .......... 87
4.7 Sensitivity analysis for the bale sequence order. ............................. 87
4.8 Inventory levels for different bale sequence order for $r = 2.95 \text{ dt/hr.}$ ............................... 88
Chapter 1

Introduction

1.1 Background

Stochastic programming is a well-known modeling and solution framework for optimization problems that involve uncertain data. These problems arise in a broad range of application areas, such as finance, healthcare, military, supply chain, and energy. In such problem settings, ignoring the uncertainty may lead to inferior or simply wrong decisions. We refer the reader to Ruszczyński and Shapiro (2003), Shapiro et al. (2009), and Birge and Louveaux (2011) for extensive discussions on stochastic programming.

Considering the adaptability and timing of decisions relative to how the uncertainty unfolds over time, we may choose different types of stochastic programming models, namely chance-constrained stochastic programs, two-stage stochastic programs, and multi-stage stochastic programs. If we are going to make a one-shot decision before observing the realization of the uncertain information, then we focus on controlling the risk associated with our decision. One common approach is to limit undesired outcomes with a risk tolerance threshold that is chosen by the decision-maker. This class of stochastic programming models is called the chance-constrained stochastic program (Charnes and Cooper, 1960). A generic chance-constrained stochastic programming model can be formulated as follows:

\[
\min \{ f(x) \mid P\{G(x, \xi) \leq 0\} \geq (1 - \epsilon), x \in \mathcal{X} \}.
\]

(1.1)

Here \( \mathcal{X} \subseteq \mathbb{R}^n \) represents the deterministic feasible region for decision variable \( x \in \mathcal{X} \), \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is a continuous function representing the objective to be minimized, and \( \xi \) is a random vector, whose probability
distribution $P$ is supported on set $\Xi \subset \mathbb{R}^d$, where $d$ is the dimension of the random data. In addition, $G : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ is a given constraint mapping and $\epsilon \in (0, 1)$ is the risk tolerance that the decision-maker sets.

If the decision-maker has the opportunity to take one additional corrective action after observing the realization of uncertain data, then the overall decision-making process can be modeled in two stages. This framework is called \textit{two-stage} stochastic programs with recourse (Beale, 1955; Dantzig, 1955). In general, the focus of these problems is the long-run average performance, optimizing decisions that will be made repeatedly in an uncertain environment. Thus, the objective function is modeled to represent the long-run expected cost/revenue. In a two-stage stochastic program, one set of decision variables $x \in X$ is determined before the decision-maker observes the realizations of random variables $\xi$, and they are referred to as the \textit{first-stage} decision variables. The other set of decision variables $y(\xi)$ are determined upon the realizations of random variables; hence, they are dependent on $\xi$. These variables are called \textit{second-stage} decision variables or the \textit{recourse variables}. A two-stage stochastic linear program (LP) can be written as follows:

$$\min_{x \in X} c^\top x + \mathbb{E}[f(x, \xi)] \quad (1.2a)$$

where

$$f(x, \xi) := \min \tilde{q}^\top y \quad (1.3a)$$

s.t. $\tilde{T}x + \tilde{W}y = \tilde{h}, \quad (1.3b)$

$y \geq 0. \quad (1.3c)$

Here $\xi := (\tilde{q}, \tilde{T}, \tilde{W}, \tilde{h})$ are the uncertain data of the second-stage problem. The expectation $\mathbb{E}[f(x, \xi)]$ is taken with respect to the probability distribution $P$.

Suppose the uncertain information is revealed over time and the decisions are made sequentially based on the data that has been realized so far in each stage. In this case, the \textit{multi-stage} stochastic programming framework is appropriate to model the nested structure of the decision-making process, which involves the interplay between decisions and uncertainty. In a planning horizon of $T$ stages, the dynamic realization of uncertainty is typically modeled as a stochastic process $(\xi_1, \xi_2, \ldots, \xi_T)$.  

2
A $T$-stage stochastic program can be written in the following general nested formulation:

$$
\min_{x_1 \in X_1(x_0, \xi_1)} f(x_1, \xi_1) + \mathbb{E}[\xi_1] \left[ \min_{x_2 \in X_2(x_1, \xi_2)} f(x_2, \xi_2) + \mathbb{E}[\xi_2] \left[ \cdots + \mathbb{E}[\xi_{T-1}] \left[ \min_{x_T \in X_T(x_{T-1}, \xi_T)} f(x_T, \xi_T) \right] \right] \right]. \quad (1.4)
$$

Here $\xi_t := (\xi_1, \xi_2, \ldots, \xi_t)$ represents the history of this stochastic process up to time $t$. The expectation $\mathbb{E}[\xi_t]$ is taken with respect to the conditional probability measure of the random vector $\xi_{t+1}$.

Unfortunately, formulation (1.4) is computationally challenging due to the nested optimization posed by the sequential nature of the decision-making structure. This challenge can be addressed by using a dynamic programming reformulation. Under the following stage-wise independence assumption, dynamic programming formulation can be solved efficiently using the Stochastic Dual Dynamic Programming (SDDP) algorithm proposed by Pereira and Pinto (1991).

**Stage-wise independence assumption** considers that $\xi_t$ is independent of the history of the stochastic process up to time $t-1$, for $t = 1, 2, \ldots, T$, which is given by $\xi_{t-1}$.

Under this assumption, the multi-stage problem (1.4) can be written in the following generic dynamic programming formulation:

$$
\min_{x_1 \in X_1(x_0, \xi_1)} f(x_1, \xi_1) + Q_2(x_2), \quad (1.5)
$$

where $Q_{t+1}(x_t) := \mathbb{E}[Q_{t+1}(I_t, m_{t+1})], \forall t \neq T$, is referred to as the expected cost-to-go function. Given any realization of the random vector $\xi_t$ and $x_{t-1}$, the optimization problem to be solved in stage $t$ is given by:

$$
\min_{x_t \in X_t(x_{t-1}, \xi_t)} f(x_t, \xi_t) + Q_{t+1}(x_t), \quad (1.6)
$$

and $Q_{T+1}(x_T) = 0$.

A fundamental assumption in stochastic programs is that the random variables follow a known probability distribution. In most cases, calculating the expected value $\mathbb{E}[f(x, \xi)]$ is computationally challenging since it requires high-dimensional integration. To address this challenge, a commonly used approach is Sample Average Approximation (SAA). In SAA we assume samples can be drawn from the probability distribution of $\xi$. Using these independent and identically (i.i.d.) drawn samples, which are called scenarios in stochastic programming, we can construct a discrete probability distribution with finite support to approximate the true distribution.

Given a set of scenarios $\{\xi_s\}_{s \in \mathcal{N}}$ that is generated from a Monte Carlo i.i.d. sample, where $\mathcal{N}$ is the
index set of these scenarios, the expectation $\mathbb{E}[f(x,\xi)]$ in (1.2a) is approximated by:

$$
\frac{1}{|N|} \sum_{s \in N} f(x,\xi_s),
$$

and the following is the resulting SAA problem of the two-stage SP (1.2):

$$
\begin{align*}
\min & \ c^T x + \frac{1}{|N|} \sum_{s \in N} q_s^T y_s \\
\text{s.t.} & \ T_s x + W_s y_s = h_s, \ \forall s \in \mathcal{N}, \\
& \ y_s \geq 0, \ \forall s \in \mathcal{N}, \\
& \ x \in X.
\end{align*}
$$

Under mild regularity conditions, the optimal objective value $\hat{v}_N$ and an optimal solution $\hat{x}_N$ of the SAA problem (1.7) converge to the optimal objective value $v^*$ and an optimal solution $x^*$ of the true problem (1.2) w.p. 1, as $|N| \to \infty$ (Shapiro, 2003). Mak et al. (1999) show that, in expectation, $\hat{v}_N$ is a lower bound on $v^*$ and as $|N|$ increases this bound improves. We refer the reader to (Pagnoncelli et al., 2009) for a detailed discussion of SAA of chance-constrained stochastic programs, and to (Shapiro, 2006) for a detailed discussion of SAA of multi-stage stochastic programs.

In this dissertation, we focus on practical problems that arise in the bioenergy sector. The increasing interest to generate clean, affordable, and secure energy has led the energy sector to renewable energy resources (McCrone et al., 2017). Renewable energy resources (e.g., solar, wind, bioenergy, and hydropower) are resources that will not deplete or can be naturally replenished. During the last decade, countries worldwide invested $2.7 trillion in renewable energy resources collectively (UN Environment Programme et al., 2020). The 2020 United Nations Environment Programme report states that this investment is more than three times the amount invested over the years 2000-2009 (UN Environment Programme et al., 2020).

The key appealing advantage of renewable energy resources is their environmental benefits. Renewable energy systems are generally referred to as “green” or “clean” since they generate little waste and pollutants, if any. Additionally, waste-to-energy technologies can even convert wastes into biofuel and contribute to the solution of the waste disposal problem. In general, renewable energy resources have much lower greenhouse gas (GHG) emissions compared to the conventional fossil fuel systems (Amponsah et al., 2014). The International Renewable Energy Agency (IRENA) published their proposed pathway to (i) limit the global temperature rise to 1.5°C and (ii) bringing CO$_2$ emissions closer to net-zero by 2050 (Gielen et al., 2014).
According to the IRENA’s pathway, over 90% of the energy solutions by 2050 will involve renewable energy (Gielen et al., 2021).

Renewable energy resources present great potential for the energy sector and could contribute to the sustainability of the world to a great extent. However, they have their own set of unique challenges, which limits the reliability and cost-efficiency of renewable energy systems. As a result, we as humankind face difficulties in preventing global warming and decreasing GHG emissions (Gielen et al., 2021). The variation in biomass feedstock characteristics results in high costs both in supply chain management and biorefinery operations.

Below we provide a summary of decision-making problems arising in the bioenergy sector which are addressed in this dissertation:

- **Biomass supply chain design and optimization**: These problems consider strategic and tactical decisions pertaining to the location, production, inventory, and distribution in biomass supply chains. High-level decisions in these problems impact the availability, the quality, and the cost of biomass. Questions, such as where to locate the plant(s), what size of a plant should be constructed, which supplier(s) to select, and what kind of transportation mode to choose, need to be addressed to strengthen the sustainable bioeconomies, enhance the security of energy supplies, and reduce GHG emissions.

- **Biomass blending**: Quality indicators for biomass include moisture, thermal, and ash content which impact the conversion rates and system efficiencies. Typically, high-quality biomass feedstocks are expensive (Perez-Verdin et al., 2009). Blending biomass feedstock of different physical or chemical properties provides an opportunity to adjust feedstock quality in order to meet annual demand of the biomass conversion plants at minimum cost. Some of the key questions that should be addressed in this context are: (i) Does biomass blending practice impact reducing the cost of producing biofuels?, and (ii) What should be the biomass mixing ratios?.

- **Optimization of biorefinery operations**: Biorefinery operations can be categorized as biomass preprocessing, biomass handling and biomass storage. Recent years have witnessed that without these components efficiently operate, biofuel industry is unable to compete with fossil fuels even with significant investments to the biomass supply chains. There is a need for models to streamline processes within a biorefinery and identify the optimal equipment control policies. It is also necessary to study strategies such as biomass bale sequencing and inventory holding so as to improve the reliability and utilization of the biorefinery equipment.
1.2 Overview of the Dissertation

In Chapter 2, we study a supply chain design problem. Specifically, we introduce the challenges seen in biomass supply chains due to the variations in biomass characteristics. We consider a supply chain with a number of suppliers who provide biomass feedstock to a single biorefinery which uses a thermochemical conversion process to produce biofuel. We consider the blending of biomass feedstocks that have different physical or chemical properties. This practice provides an opportunity to meet the specifications of the conversion platform in a cost-effective way through adjusting the feedstock mix. The conversion processes require that the ash content of the biomass blend used must be lower than a user-defined threshold value, and the total thermal content must be higher than another threshold value. Meeting these requirements is a challenge because biomass ash and thermal contents are random and vary by supplier. Two critical aspects of this problem shaped our modeling choice. First, the supplier and biomass feedstock selections are long-term strategic decisions that are made before observing uncertain biomass characteristics. Second, the quality requirements of the conversion process are soft requirements. In other words, the biorefinery is willing to accept a particular risk level regarding these quality requirements as long as the supply chain costs are decreased significantly. Considering these key problem structures, we propose chance-constrained stochastic programming models and solution methods. We utilize the sample average approximation (SAA) method to approximate the chance constraints and propose solution algorithms to solve this approximation. We develop a case study for South Carolina using data provided by the Billion Ton Study (Langholtz et al., 2016). This case study enables us to show the quality of proposed algorithms and provide managerial insights to biorefinery management for biomass blending ratios.

In Chapter 3, we shift our focus to more operational challenges and study biorefinery operations. Inside the biorefinery, variations in physical and chemical characteristics of biomass could lead to an uneven flow of biomass, which reduces equipment utilization and increases operational costs. Consequently, the system reliability could be challenged. We define the system reliability as the probability of achieving the desired target feeding rate under uncertain biomass moisture content, particle size distribution, and equipment failure. Chance-constrained stochastic programming provides an effective decision-making framework to model the reliability requirement of a system. In addition, adjusting the risk parameter $\epsilon$ facilitates the analysis of the trade-off between reliability and costs. We propose a chance-constrained stochastic programming model, which identifies operating conditions of equipment and inventory level to maintain a continuous biomass flow to the reactor. We conduct an extensive computational analysis on the impacts of sequencing of biomass
bales, storage capacities, and particle size on operational costs and reactor utilization.

There are recent efforts in manufacturing facilities to integrate physical assets (e.g., vehicles, equipment, machines, robots), people, and information systems. Integration of these different elements will result in higher productivity, better quality control, and more flexible operations. In light of this trend, we study biorefinery operations further. In recent years, biomass processing facilities have incorporated sensors to monitor the changes in biomass characteristics. Sensors reveal the uncertain biomass moisture content over time and enable the operators to revise their operational decisions according to the changes in biomass moisture. In Chapter 4, we develop a novel multi-stage stochastic programming model that can capture this sequential structure and integrate the monitoring technologies with the decision-making process. Considering the computational challenges of multi-stage stochastic programs, we model the system reliability with a penalization approach. We ensure that a prescribed utilization target of the reactor is consistently achieved by penalizing the violation of the target reactor feeding rate. A case study is developed using data collected at Idaho National Laboratory’s biomass processing facility. We show the value of multi-stage stochastic programming from an extensive computational experiment and provide managerial insights on biomass processing operation based on our sensitivity analysis.

Lastly, we present the conclusion and a summary of this dissertation in Chapter 6.
Chapter 2

A Stochastic Biomass Blending Problem in Decentralized Supply Chains

2.1 Background

The majority of the existing biomass logistics models are focused on reducing the total costs of delivering biomass to conversion plants. This is mainly because the emerging biomass supply system inherited models (and the underlying assumptions) from the existing agricultural and logging industries. These models pay little attention to the impacts that biomass quality has on costs and conversion rates (Kenney et al., 2013).

Typically, feedstocks of high quality are expensive. Blending biomass feedstock of different physical or chemical properties provides an opportunity to adjust the quality of the feedstock to meet annual needs of the conversion platform at the minimum cost. Quality indicators for biomass are moisture content, thermal content, ash content, etc. For example, clean pine is frequently identified as the biomass feedstock of choice for Thermochemical conversion process. Factors which impact the performance of pyrolysis are oxygen and hydrogen content of biomass. Blending clean pine in appropriate proportions with logging residues results in a blend that maintains the desired levels of ash content, while using relatively less expensive materials. Blending for such purposes is a common practice in many industries (Hill, 1990). Other examples are blending...
of animal feed to obtain specific nutrient requirements (Reddy et al., 2009); and blending of high-ash biomass with low-ash coal to allow their use for biopower generation (Sami et al., 2001).

We propose a stochastic optimization model that identifies a blending of different types of biomass to meet process requirements of a biorefinery at minimum cost. Current processes require that the ash content of the blend used must be lower than a threshold value, and total thermal content must be higher than another threshold value. Meeting these requirements is a challenge because ash and thermal contents of biomass are random and vary by supplier. These process requirements are soft since, for example, if ash content is higher than the threshold value, the blend undergoes preprocessing which reduces ash content. Similarly, if thermal content is low, then additional biomass can be purchased to meet the thermal requirement by contracting new suppliers. However, these practices are expensive. Thus, rather than meeting ash and thermal requirements all the time, which imposes high costs to the system, in rare occasions, violation of these requirements should be allowed (e.g., 10-20% of the time). We model these soft requirements using chance constraints.

The U.S. Department of Energy (DOE) and other federal agencies have made significant investments to help the bioenergy industry grow. Despite these investments, this industry remains a nascent concern and unable to compete with fossil fuels. The purpose of this work is to evaluate the potential impact that practices, such as biomass blending, have to reduce the cost of producing biofuels. To achieve this, we propose two supply chain models, one that assumes centralized decision making in which a single decision maker has full control, and another model that assumes decentralized decision making in a noncooperative environment. The decentralized model considers that the biorefinery and suppliers are independent entities who have their own goals and objectives. In this model, each entity makes decisions to improve its own performance rather than the performance of the overall supply chain. To address this, we model the relationships between the biorefinery and suppliers via a Stackelberg game.

The objective of comparing the centralized and decentralized decision making is two-fold. First, we use these models to estimate the impact of collaboration among supply chain members to costs, suppliers selection and biomass purchasing decisions. Second, we estimate the approximation error from assuming centralized decision making in the supply chain. Decision makers can use these results to (i) motivate potential collaborations in the supply chain; (ii) evaluate the trade-offs between model accuracy and solution time.

We develop a case study focused on South Carolina (SC). To develop this case study we used the Billion Ton Study (Langholtz et al., 2016) which provides data about the availability and cost of different types of biomass in each county of SC. This cost includes the cost of land, and the cost of planting, harvesting,
and collecting biomass. We also used the Bioenergy Feedstock Library (INL, 2017) which is developed and maintained by the Idaho National Laboratory (INL) with sponsorship from the DOE. This library is both a physical repository and knowledge database of biomass feedstock which provided the data about ash and thermal content of biomass feedstocks. We use this case study to validate the models proposed and conduct numerical experiments.

2.2 Literature Review

The research presented in this chapter is related to the following three main streams of literature: biomass blending problem, chance constraint programming, and bilevel optimization.

**Biomass Blending Problem**: This problem is a special case of the blending problem, which is well studied by the operations research community. It refers to models that aid a “decision concerning the best use of an organization’s resources when mixing or blending various ingredients to produce a final product without violating certain specified criteria.” There is a variety of applications of this problem ranging from the oil and coal industries (Shih and Frey, 1995; Sami et al., 2001), to the chemical industry (Urquhart et al., 2004; Zhengping et al., 2004), to food industry (Stigler, 1945; Hashempour-Baltork et al., 2016), etc.

The research on biomass blending is scarce. The existing literature is mainly confined to the study of the impact that biomass blending has on the conversion performance (Shi et al., 2013). These studies use sensitivity analysis to capture the impacts of biomass supply/quality on conversion rate (Jacobson et al., 2014a). Literature exists about mathematical models that focus on optimizing blending of coal and grain products. For example, (Sivaraman et al., 2002) propose a model that identifies blending ratios to maximize revenues from sales of blended grain products. Work by (Shih and Frey, 1995) proposes a multi-objective optimization model to identify coal blends which minimize system wide costs and GHG emissions. To the best of our knowledge, there are no papers which focus on the impacts of biomass blending on supply chain costs.

**Chance Constraint Programming**: Many applications in supply chain (Lejeune and Ruszczynski, 2007), production planning (Murr and Prekopa, 2000), energy systems (Wang et al., 2012), etc. use chance constraint programming (CCP) to model uncertainties. These models ensure that the probability of meeting a requirement is above a certain threshold level. These models are typically very difficult to solve (Birge and Louveaux, 2011) for the following two reasons. First, the probability of meeting a certain constraint cannot be computed exactly due to the computational challenge of multidimensional integration. Second, the
feasibility region defined by the chance constraints may not be convex (Kim et al., 2015). The computational
difficulties for solving CCPs motivated the development of approximate solution approaches. There are
two main approaches to solve CCP models. The first approach discretizes the corresponding probability
distribution and solves the corresponding combinatorial problem (Dentcheva et al., 2000; Luedtke and Ahmed,
2008). The second approach develops convex approximations of the chance constraints (Nemirovski and
Shapiro, 2006).

This research uses the SAA method. The approximation is obtained by replacing the actual distribu-
tion in a chance constraint by an empirical distribution corresponding to a random sample. The resulting
deterministic equivalent model ensures that the number of unexpected “failures” in these independent samples
is below the model thresholds. Variations of the SAA for chance constrained problems have been investigated
in (Atlason et al., 2008; Luedtke and Ahmed, 2008). The theoretical properties of SAA have been studied in
(Pagnoncelli et al., 2009). These studies provide the conditions for which an upper and a lower bound to the
original CCP problem can be obtained.

Bilevel Optimization: A bilevel optimization model is a mathematical model with an optimization
problem in the constraints. This model is a generalization of the Stackelberg game. Many applications in re-
source planning, financial planning, land-use planning, etc. are modeled and solved using bilevel optimization
models (Lu et al., 2006). Bilevel optimization models are difficult to solve since the corresponding feasible
region is not convex. A special case is the bilevel optimization model where the inner optimization model
is linear. In this case, the inner optimization model can be replaced by the corresponding KKT conditions.
The corresponding single-level problem is a nonlinear program. A number of approaches have been de-
veloped to solve these nonlinear programs (Bard, 1984, 1998; Moore and Bard, 1990). In addition to the KKT
based approaches, other solution methods have been developed, such as, descent methods, penalty function
methods, and trust-region methods (Sinha et al., 2017).

Within the framework of any bilevel optimization, a leader’s decision is influenced by the reaction of
his follower(s). In a setting with multiple followers, a leader’s decision is influenced not only by the decision
of each follower, but also by the relationships among these followers. Work by (Lu et al., 2006) identifies nine
different kinds of relationships amongst followers by establishing a general framework for bilevel multi-follower
decision problems. One of the problems analyzed in great detail is the uncooperative decision problem, which
is the same problem we solve in this research. (Lu et al., 2006) extend the Kuhn–Tucker approach to find an
optimal solution for the uncooperative decision model and illustrate its performance via a real life case study.
2.3 Blending Problem in a Centralized Supply Chain

2.3.1 Assumptions and Parameters

Consider a supply chain with $|I|$ suppliers who provide biomass feedstock to a single biorefinery which uses Thermochemical conversion process to produce biofuel. Figure 2.1 provides a graphical representation of the design of this supply chain, and the Thermochemical conversion process. Biomass qualities (such as ash and thermal contents) influence the effectiveness of the Thermochemical process (Tanger et al., 2013).

The following parameters describe the biorefinery and are a function of its production capacity: $\alpha$ denotes the allowable ash content (in %) and $\tau$ denotes the annual thermal requirement (in British Thermal Units (BTU)). Let $S_{ib}$ denote the amount of biomass feedstock $b$ available at supplier $i$. Let $\tilde{a}_{ib}$ denote the corresponding ash content, and $\tilde{h}_{ib}$ denote the thermal content. We assume that ash and thermal contents are stochastic and are represented by random variables which follow some continuous distributions.

Let $c_p$ correspond to the unit cost of biomass at the farm gate. This cost includes the cost of harvesting and collecting biomass. The Billion Ton Study conducted by the Oak Ridge National Laboratory Langholtz et al. (2016) shows that this cost depends on the amount of biomass available, and provides data about the amount of biomass available for a given set of farmgate costs for each county in the USA (see Figure 2.2a). Let $\mathcal{P} = \{10, 20, \ldots\}$ be the set of unit farmgate costs listed in the Billion Ton Study. We use $p = 1, \ldots, |\mathcal{P}|$ to denote the index and $c_p$ is an element of this set. Let $k_{ibp}$ denote the least and $\bar{k}_{ibp}$ denote the maximum amount of biomass available at cost $c_p \in \mathcal{P}$, thus, $S_{ib} = \bar{k}_{ibp}$. The relationship between the amount available and the total purchasing cost at a biorefinery is represented via a piece-wise linear function (see Figure 2.2b).

Another problem parameter is $t_{ib}(= v_b \times Dist_i + g_b)$ which denotes the unit transportation cost (in $$/ton). This cost depends on the type of biomass delivered and the distance from the supplier $i$, $Dist_i$; the variable unit cost, $v_b$; and the fixed unit transportation cost, $g_b$. Finally, $f_b$ denotes the cost of processing
and holding inventory at the biorefinery.

![Graphs showing supply curve and total purchasing cost](image)

**Figure 2.2:** Supply Curve for Woody Biomass in Pickens County, SC.

We provide a summary of the notation used in Appendix A.5.

### 2.3.2 A Stochastic Problem Formulation

Let $X_{ib}$ denote the amount of biomass type $b$ purchased from supplier $i$. Then, the corresponding biomass purchase cost is represented by:

$$
F_{ib}(X_{ib}) = \begin{cases} 
    c_1 X_{ib}, & \text{if } 0 \leq X_{ib} \leq \bar{k}_{ib1}, \\
    c_2 X_{ib}, & \text{if } \bar{k}_{ib2} < X_{ib} \leq \bar{k}_{ib2}, \\
    \vdots, & \\
    c_{|P|} X_{ib}, & \text{if } \bar{k}_{ib|P|} < X_{ib} \leq \bar{k}_{ib|P|},
\end{cases}
$$

(2.1)

where, $\bar{k}_{ibp} = \frac{k_{ib(p+1)}}{k_{ib(p+1)}}$ for $p = 1, \ldots, |P| - 1$.

The proposed model identifies a blendstock that minimizes the total supply chain costs including biomass purchasing, transportation, and processing and inventory costs at the biorefinery.

$$
\min : \sum_{i \in I} \sum_{b \in B} \left[ F_{ib}(X_{ib}) + t_{ib} X_{ib} + f_b X_{ib} \right].
$$

(2.2)

This objective is minimized subject to the following constraints. Constraints (2.3) indicate that the
amount of biomass shipped from supplier \(i\) is limited by its availability.

\[
X_{ib} \leq S_{ib} \quad \forall i \in I, b \in B. \tag{2.3}
\]

Since biomass quality impacts the performance of the conversion process, biorefineries require that the total ash content be at most \(\alpha\%\) of biomass purchased, and the corresponding total thermal value be at least \(\tau\) BTUs. However, these are soft requirements. That means the biorefinery would like to meet these requirements. However, on a few occasions, the biorefinery is willing to violate these requirements if doing so will sufficiently decrease supply chain costs.

Let \(\beta\) and \(\gamma\) represent the risk parameters chosen by the biorefinery. These values are typically 10\% to 20\%. The following chance (probabilistic) constraint indicates that ash content of biomass purchased by the biorefinery should be smaller than a threshold level \(\alpha\) at least \((1 - \beta)\%\) of the time.

\[
Pr \left( \sum_{i \in I} \sum_{b \in B} \tilde{a}_{ib}X_{ib} \leq \alpha \sum_{i \in I} \sum_{b \in B} X_{ib} \right) \geq 1 - \beta. \tag{2.4}
\]

Similarly, the following chance constraint indicates that the thermal content of biomass purchased by the biorefinery should be greater than the threshold level \(\tau\) at least \((1 - \gamma)\%\) of the time. Note that the thermal energy gathered from the biomass delivered depends on its thermal content \((\tilde{h}_{ib})\) and efficiency \((e_b)\) of the conversion process.

\[
Pr \left( \sum_{i \in I} \sum_{b \in B} e_b \tilde{h}_{ib}X_{ib} \geq \tau \right) \geq 1 - \gamma. \tag{2.5}
\]

Constraints (2.6) are the non-negativity constraints.

\[
X_{ib} \geq 0 \quad \forall i \in I, b \in B. \tag{2.6}
\]

Formulation (2.2) to (2.6) is a chance constraint programming (CCP) model. We call this model (\(P\)).

### 2.4 Solution Approaches: Centralized Blending Model

Initially, we present a mixed-integer programming (MIP) formulation of model (\(P\)). Next, we propose a sample average approximation (SAA) of chance constraints (2.4) and (2.5). Finally, we present a linear approximation of \(F_{ib}(X_{ib})\) and a corresponding MIP formulation.
2.4.1 A Mixed-Integer Formulation of \((P)\)

Let \(X_{ibp}\) be a decision variable which presents the amount of biomass type \(b\) purchased from supplier \(i\) which falls in bracket \(p\). Let \(Z_{ibp}\) be a binary variable which takes the value 1 when the amount of biomass type \(b\) purchased from supplier \(i\) falls in bracket \(p\), and takes the value 0 otherwise.

The following is a mixed-integer formulation of \((P)\).

\[
\begin{align*}
(\bar{P}) : \quad & \min \sum_{i \in I} \sum_{b \in B} \sum_{p \in \mathcal{P}} c_{ibp} X_{ibp} \\
\text{s.t.} \quad & Pr \left( \sum_{i \in I} \sum_{b \in B} \sum_{p \in \mathcal{P}} (a_{ib} - \alpha) X_{ibp} \leq 0 \right) \geq 1 - \beta, \quad (2.7b) \\
& Pr \left( \sum_{i \in I} \sum_{b \in B} \sum_{p \in \mathcal{P}} e_{b} h_{ib} X_{ibp} \geq \tau \right) \geq 1 - \gamma, \quad (2.7c) \\
& \sum_{p \in \mathcal{P}} X_{ibp} \leq S_{ib}, \quad \forall i \in I, b \in B, \quad (2.7d) \\
& k_{ibp} Z_{ibp} \leq X_{ibp} \leq k_{ibp} Z_{ibp}, \quad \forall b \in B, i \in I, p \in \mathcal{P}, \quad (2.7e) \\
& \sum_{p \in \mathcal{P}} Z_{ibp} = 1, \quad \forall i \in I, b \in B, \quad (2.7f) \\
& X_{ibp} \geq 0, \quad \forall i \in I, b \in B, p \in \mathcal{P}, \quad (2.7g) \\
& Z_{ibp} \in \{0, 1\}, \quad \forall i \in I, b \in B, p \in \mathcal{P}, \quad (2.7h)
\end{align*}
\]

where, \(c_{ibp} = c_p + t_{ib} + f_b\). Notice that, constraints (2.7d) are redundant because constraints (2.7e) and (2.7f), and the definition of \(k_{ibp} = S_{ib}\) impose the same restrictions on the solutions of \((P)\).

2.4.2 A Sample Approximation of Chance Constraints

In order to make it easier for the reader to follow the description of the SAA, we provide the following succinct formulation of \((\bar{P})\).
\[ (\tilde{P}) : \quad \min_{x \in \mathcal{Z}} Z = f(x) \]

\[ p^1(x, \tilde{a}) \leq \beta, \]

\[ p^2(x, \tilde{h}) \leq \gamma. \]

Let \( E^1(x, \tilde{a}) = \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} (\tilde{a}_{ib} - \alpha) X_{ibp} \), and, \( p^1(x, \tilde{a}) = \Pr \left( E^1(x, \tilde{a}) > 0 \right) \) (note: \( \Pr \left( E^1(x, \tilde{a}) > 0 \right) \leq \beta \), is equivalent to \( \Pr \left( E^1(x, \tilde{a}) \leq 0 \right) \geq 1 - \beta \). Let \( E^2(x, \tilde{h}) = \tau - \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} (\epsilon_b \tilde{h}_{ib} X_{ibp}) \), and \( p^2(x, \tilde{h}) = \Pr \left( E^2(x, \tilde{h}) > 0 \right) \). Let \( \vartheta \) and \( \mathcal{X} \) represent the optimal objective function value and the feasible region of \((\tilde{P})\), respectively. Let \( \mathcal{Z} \) represents solutions which satisfy constraints (2.7d) to (2.7h). We assume that, (i) \( \mathcal{X} \) is non-empty, and (ii) \( \vartheta \) is bounded.

The literature uses triangular and uniform distributions to model \( \tilde{a}_{ib} \) and \( \tilde{h}_{ib} \), which are the random parameter in \((\tilde{P})\) (Shabani and Sowlati, 2016). Thus, \( E^1(x, \tilde{a}) \) is the linear combination of \( n_1 \) (\( n_1 = |I| \times |B| \times |P| \)) triangular distributed random variables; and \( E^2(x, \tilde{h}) \) is the linear combination of \( n_2 \) (\( n_2 = |B| \times |P| \)) uniform distributed random variables. We use Monte Carlo simulation to generate \( N \) random samples of \( \tilde{a}_{ib} \) and \( \tilde{h}_{ib} \) from the corresponding distributions and use these values to calculate \( E^1(x, \tilde{a}), E^2(x, \tilde{h}) \).

Let \( P_N^1(x) = N^{-1} \sum_{s=1}^{N} \Delta(E^1_s(x, a_s)) \) denote an empirical measure of the probability distribution of \( E^1(x, \tilde{a}) \) and \( P_N^2(x) = N^{-1} \sum_{s=1}^{N} \Delta(E^2_s(x, h_s)) \) denote an empirical measure of the probability distribution of \( E^2_s(x, \tilde{h}) \). Here, \( a_s, h_s, s = 1, \ldots, N \) are \( N \) iid realizations of \( \tilde{a}_{ib}, \tilde{h}_{ib} \), \( \Delta() \) is a measure of probability mass function value, and \( 1/N \) is the probability assigned to each realization of \( \tilde{a}_{ib}, \tilde{h}_{ib} \). The SAA replaces the original distributions of \( E^1(x, \tilde{a}), E^2(x, \tilde{h}) \) with \( P_N^1(x), P_N^2(x) \) respectively (Pagnoncelli et al., 2009).

Let \( \mathbb{I}_{(0,\infty)}: \mathcal{R} \to \{0,1\} \) be the indicator function of \((0,\infty)\), i.e.

\[
\mathbb{I}_{(0,\infty)}(t) := \begin{cases} 
1 & \text{if } t > 0, \\
0 & \text{if } t \leq 0.
\end{cases}
\]

We can now approximate \( p^1(x, \tilde{a}) \) and \( p^2(x, \tilde{h}) \) using the empirical measures \( P_N^1 \) and \( P_N^2 \) as follows

\[
\hat{p}_N^1(x) = \mathbb{E}_{P_N^1} \left[ \mathbb{I}_{(0,\infty)}(E^1(x, \tilde{a})) \right] = \frac{1}{N} \sum_{s=1}^{N} \mathbb{I}_{(0,\infty)}(E^1(x, a_s)), \tag{2.8}
\]

16
\[ \hat{p}_N^2(x) = \mathbb{E}_{P_N} \left[ \mathbb{1}_{(0,\infty)}(E^2(x, h)) \right] = \frac{1}{N} \sum_{s=1}^{N} \mathbb{1}_{(0,\infty)}(E^2(x, h_s)). \] 

Approximation \( \hat{p}_N^1(x) \) returns what proportion of times \( E^1(x, \bar{a}) < 0 \), and \( \hat{p}_N^2(x) \) returns what proportion of times \( E^2(x, \bar{h}) < 0 \). The resulting SAA approximation model is presented below.

\[ (\hat{P}) : \min_{x \in \mathcal{Z}} Z = f(x) \]
\[ \hat{p}_N^1(x) \leq \hat{\beta}, \]
\[ \hat{p}_N^2(x) \leq \hat{\gamma}. \] 

Let \( \bar{\vartheta}^N \) and \( \bar{\mathcal{X}}^N \) represent the optimal objective function value and the feasible region of \((\hat{P})\), respectively. In this formulation, the reliability levels \( 1 - \hat{\beta} \) (for \( \hat{\beta} > 0 \)) and \( 1 - \hat{\gamma} \) (for \( \hat{\gamma} > 0 \)) are different from the reliability level \( 1 - \beta \) and \( 1 - \gamma \) of the true model \((P)\). Based on Theorem 5 in (Luedtke and Ahmed, 2008), if \( \hat{\beta} < \beta \) and \( \hat{\gamma} < \gamma \), every feasible solution of \((\hat{P})\) is feasible to \((P)\) with high probability as \( N \) gets large. That is:

\[ \vartheta^N \rightarrow \vartheta \text{ and } \mathcal{X}^N \rightarrow \mathcal{X} \text{ w.p.1 as } N \rightarrow \infty. \]

Constraints (2.10) and (2.11) of \((\hat{P})\) use indicator functions. Since commercial solvers cannot handle such functions, we reformulate these constraints by introducing the following continuous variables \( \mathcal{V}, \mathcal{W}, \mathcal{U}, \mathcal{J} \) which quantify the violation of these constraints. The following are the equivalent linear constraints.

\[ E^1(x, a_s) + \mathcal{V}_s - \mathcal{W}_s = 0, \forall s = 1, \ldots, N, \] 
\[ E^2(x, h_s) + \mathcal{U}_s - \mathcal{J}_s = 0, \forall s = 1, \ldots, N, \] 
\[ \mathcal{V}_s, \mathcal{W}_s, \mathcal{U}_s, \mathcal{J}_s \geq 0, \forall s = 1, \ldots, N. \]

This reformulation minimizes the cost of violating the chance constraints (Charnes et al., 1955; Abdelaziz et al., 2007; Abdelaziz, 2012). Thus, variables \( \mathcal{W}, \mathcal{U} \) also appear in the objective function, as follows:

\[ (\bar{P}) : \min_{x \in \mathcal{Z}} Z = f(x) + \lambda \sum_{s=1}^{N} \mathcal{W}_s + \mu \sum_{s=1}^{N} \mathcal{J}_s \]
\[ \text{s.t. Constraints (2.12) – (2.14)}. \]
This problem is easier to solve as compared to \((\hat{P})\). Note however, that parameters \(\lambda\) and \(\mu\) are not known in advance. The size of these parameters is problem specific. When these penalties are too high, the minimization sets \(W_s = 0\), and \(J_s = 0\) for all \(s = 1, \ldots, N\). Consequently, \(E^1(x, a_s) \leq 0\) and \(E^2(x, h_s) \leq 0\) for all \(s = 1, \ldots, N\). Thus, we develop an algorithm which identifies the value of \(\lambda\) so that \(W_s = 0\) in at least \([(1 - \hat{\beta})N]\) of the scenarios generated; and identifies the value of \(\mu\) so that \(J_s = 0\) in at least \([(1 - \hat{\gamma})N]\) of the scenarios generated. The SAA Algorithm in Appendix A.1 is an iterative procedure which uses a binary search to identify the values of \(\lambda\) and \(\mu\).

### 2.4.3 A Linear Approximation of the Objective Function in \((P)\)

Let \(F_{ib}(X_{ib})\) be a function defined as follows:

\[
F_{ib}(X_{ib}) = \begin{cases} 
  c_1 X_{ib}, & \text{if } 0 \leq X_{ib} \leq \underline{k}_{ib1}, \\
  \lambda_{ib2} + c_2 (X_{ib} - \underline{k}_{ib2}), & \text{if } \underline{k}_{ib2} < X_{ib} \leq \overline{k}_{ib2}, \\
  \cdots & \\
  \lambda_{ib|P|} + c_{|P|} (X_{ib} - \underline{k}_{ib|P|}), & \text{if } \underline{k}_{ib|P|} < X_{ib} \leq \overline{k}_{ib|P|}.
\end{cases}
\]  

(2.15)

Where \(\lambda_{ib1} = 0\) and \(\lambda_{ibp} = \sum_{j \leq p-1} c_j (\overline{k}_{ibj} - \underline{k}_{ibj})\) for \(p = 2, \ldots, |P|\). Function \(F_{ib}(X_{ib})\) is a continuous and convex approximation of function \(F_{ib}(X_{ib})\). Furthermore, \(F_{ib}(X_{ib})\) provides an outer-approximation of \(F_{ib}(X_{ib})\).

The following is an approximations of \((P)\).

\[
(\hat{P}): \min \sum_{i \in I} \sum_{b \in B} F_{ib} + \lambda \sum_{s=1}^{N} W_s + \mu \sum_{s=1}^{N} J_s \\
\text{s.t.} \\
(2.12) - (2.14), \\
F_{ib} \geq \lambda_{ibp} + c_p (X_{ib} - \underline{k}_{ibp}), \quad \forall i \in I, b \in B, p \in P, \\
X_{ib} \geq 0, \quad \forall i \in I, b \in B.
\]  

(2.16)

**Proposition 1.** The feasible region of \((\hat{P})\) is convex. (Proof in Appendix A.2.)

**Proposition 2.** For each feasible solution of \((\hat{P})\) one can find a feasible solution of \((\hat{P})\), and vice versa. (Proof in Appendix A.2.)
Proposition 3. An optimal solution of $(\hat{\mathcal{P}})$ is a lower bound of $(\bar{\mathcal{P}})$. (Proof in Appendix A.2.)

Based on Propositions 2 and 3, we develop an Algorithm for the Centralized Problem $(\bar{\mathcal{P}})$. The algorithm solves $(\hat{\mathcal{P}})$ to obtain a feasible solution $X^*$. Based on Proposition 2, $X^*$ is feasible for $(\bar{\mathcal{P}})$ as well. Next, we find an upper bound for $(\bar{\mathcal{P}})$ by calculating its objective function value at $X^*$. In order to evaluate the quality of the approximation we find a lower bound for $(\bar{\mathcal{P}})$ by calculating the objective function value of $(\hat{\mathcal{P}})$ at $X^*$. We report the corresponding error gap.

Note that Algorithm for the Centralized Problem solves $(\bar{\mathcal{P}})$ for a given value of $\lambda$ and $\mu$. We use this algorithm within the SAA Algorithm in order to identify the best values of $\lambda$ and $\mu$ that optimize $(\bar{\mathcal{P}})$.

2.5 Blending Problem in a Decentralized Supply Chain

Due to the computational challenges of solving models for decentralized supply chains, most of the works in the literature assume centralized problem setting. This assumption makes the problem easy to solve, but often such a setting is not realistic. Centralized decision making assumes that players of a supply chain coordinate their decisions. However, this is rarely the case since each player tries to maximize own profits (Chopra et al., 2013). Lack of coordination in the supply chain leads to the bullwhip effect, increased inefficiencies and higher supply chain costs (Lee et al., 1997; Chiu and Kremer, 2013; Hassanzadeh et al., 2014). In this section we model this blending problem in a decentralized supply chain.

We consider that the biorefinery and suppliers are independent entities who have their own goals and objectives. We propose a Stackelberg game to model these relationships. The biorefinery is the leader of the game since it is typically a large enterprise. Suppliers are the followers in the game since farms in the USA are typically of small and medium size. Transportation costs in this supply chain are high because biomass is bulk product and has low energy density. Therefore, to keep transportation costs low, a biorefinery purchases from farms located nearby. As a result, we assume that one single biorefinery (the leader of the game) is located in the area.

The game begins with the biorefinery setting a “door” price based on the type of biomass supplied. The “door” price represents the amount of money paid to the supplier per ton of biomass delivered to the door of the biorefinery. The goal of the biorefinery is to identify a blendstock which minimizes its total supply chain costs while meeting thermal and ash content requirements. Next, suppliers decide how much to offer. If the amount offered meets the needs of the biorefinery, the game ends. Otherwise, the biorefinery adjusts the prices offered to suppliers, and the game continues. In this game, suppliers are independent, and each
supplier focuses on maximizing his own profits.

**Leader’s Problem:** The biorefinery leads the game by setting a “door” price. Let \( C_b \) denote this price which is charged based on biomass type. The objective of the leader is to identify a blendstock that minimizes her total supply chain costs. This objective is presented by the following equation.

\[
\min : Z^L = \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} (C_b + f_b) X_{ibp}.
\]

Since biomass quality impacts the performance of the conversion process, the biorefinery requires that ash content (2.7b), thermal content (2.7c), and non-negativity (2.7g) requirements are met. Additionally, we assume that farms will participate in this game only when it leads to profits. This assumption is realistic since, in a free market economy, farmers would not choose to participate in the game if doing so leads to an economic loss. Let \( \tilde{c}_{bp} \) represent the cost of harvesting, collecting, and storing biomass \( b \) within the cost bracket \( p \in P \). \( t_{ib} \) is the unit transportation cost. We consider that each farm faces the same cost brackets which are defined in the centralized model \((E_{ibp}, \bar{k}_{ibp})\). Thus, the following is a constraint of leader’s problem.

\[
\sum_{b \in B} \sum_{p \in P} (C_b - \tilde{c}_{bp} - t_{ib}) X_{ibp} \geq 0, \quad \forall i \in I.
\]

Finally, \( C_b \geq 0, \quad b \in B \). \( (2.20) \)

Let \( X^L \) denote the feasible region defined by constraints (2.7b), (2.7c), (2.7g), (2.19), (2.20); and let the corresponding model be the *leader’s problem* \((P^L)\).

**Followers’ Problem:** The objective of each farm is to identify how much biomass to supply in order to maximize profits. The objective of farm \( i \in I \) is the following.

\[
\max : Z^F_i = \sum_{b \in B} \sum_{p \in P} (C_b - \tilde{c}_{bp} - t_{ib}) X_{ibp}.
\]

The amount of biomass delivered by a farm is impacted by its availability and the price offered by the biorefinery. Constraints (2.7e) to (2.7h) represent these restrictions.

Let \( X^F_i \) denote the feasible region of the problem faced by follower \( i \) defined by (2.7e) to (2.7h). Let the corresponding model be the *followers’ problem* \((Q^F_i)\) and \( Z^F \) represent the set of optimal solutions. A complete formulation of the proposed bilevel optimization model is presented in Appendix A.3. We refer to this as formulation \((Q)\). Next, we provide the corresponding succinct formulation in order to make it easier
for the reader to follow the approach we develop for solving the bilevel optimization model proposed.

\[
(Q) : \min_{x^*, c^l} Z^L = F(x^*, c^l) \\
(x^*, c^l) \in \mathcal{X}^L \\
x^* \in \arg \max_{x^f, z^f} Z^F = f_i(x^f_i, z^f_i) \quad \forall i \in I \\
(x^f_i, z^f_i) \in \mathcal{X}_i^F.
\]

Let \((\bar{Q})\) represent the SAA of model \((Q)\) where the probabilistic constraints (2.7b) and (2.7c) are linearized as described in Section 2.4.2. A complete formulation of \((\bar{Q})\) is presented in Appendix A.3.

### 2.6 Solution Approaches: Decentralized Blending Model

We consider the optimistic (or strong) formulation of \((Q)\). As such, the farms select from their set of optimal solutions, the amount of biomass to supply according to what is best for the biorefinery.

In this section we focus on solving \((\bar{Q})\). Note that, \((\bar{Q})\) is not convex due to the inner optimization model (the followers’ problem). Also, \((\bar{Q})\) is nonlinear due to the bilinear terms \(C_bX_{ibp}\) in the objective. Next we present an exact solution, a heuristic and a lower bound approach.

#### 2.6.1 An Exact Solution Approach to Solve \((\bar{Q})\)

##### 2.6.1.1 Analyzing Followers’ Problem:

Given the prices set by the leader \((C_b)\), the followers’ problem becomes an integer linear program. This problem can be decomposed by supplier and biomass type into \(|I| \times |E| \) sub-problems of the following
\[ \text{(} \bar{Q}_{ib}^F \text{)} : \quad \max \ Z^F = \sum_{p \in P} \bar{c} X_p, \]

s.t.

\[ \sum_{p \in P} X_p \leq S, \quad \text{(2.21a)} \]

\[ k_p Z_p \leq X_p \leq \bar{k}_p Z_p, \quad \forall p \in P, \quad \text{(2.21b)} \]

\[ \sum_{p \in P} Z_p = 1, \quad \text{(2.21c)} \]

\[ X_p \geq 0, \quad \forall p \in P, \quad \text{(2.21d)} \]

\[ Z_p \in \{0, 1\} \quad \forall p \in P. \quad \text{(2.21e)} \]

Where, \( \bar{c} = (C_b - \bar{c}_bp - t_{ib}) \).

**Proposition 4.** The linear relaxation of \( (\bar{Q}_{ib}^F) \) provides an exact solution. (Proof in Appendix A.2.)

**Lemma 1.** In an optimal solution to problem \( (\bar{Q}_{ib}^F) \), \( \bar{c} \geq 0 \) (due to constraints (2.19)). Thus, at most one \( X_p > 0 \) for \( p \in P \). Let \( p^\star \) be the cost bracket for which \( X_p > 0 \), then, \( X_{p^\star} = \bar{k}_{p^\star} \) and \( X_p = 0 \) for \( p \in P \setminus p^\star \). (Proof in Appendix A.2.)

**Theorem 1.** There is an \( O(IBP) \) algorithm which finds an optimal solution to the followers' problem \( (\bar{Q}_{ib}^F) \).

**Proof:** For each supplier and biomass type, one can identify the cost bracket which results in the maximum profits by following this two-step procedure: (1) find \( p^\star = \arg\max_{p \in P} \bar{c} \bar{k}_p \), (2) if \( \bar{c} \bar{k}_{p^\star} < 0 \), then, \( Z_p = X_p = 0 \) for all \( p \in P \), (3) if \( \bar{c} \bar{k}_{p^\star} \geq 0 \), \( X_{p^\star} = \bar{k}_{p^\star}, \ Z_{p^\star} = 1, \) and \( Z_p = X_p = 0 \) for \( p \in P \setminus p^\star \). We set \( X_{p^\star} = \bar{k}_{p^\star}, \ Z_{p^\star} = 1 \) when \( \bar{c} \bar{k}_{p^\star} = 0 \) since we consider the optimistic formulation of \( (P) \). We call this procedure **Followers Algorithm.**

This two-step procedure finds an optimal solution to \( (\bar{Q}_{ib}^F) \) in \( O(P) \), which is the time it takes to find \( p^\star \) which leads to maximum profits. Since this procedure is followed by each supplier and for each biomass type, the total running time is \( O(IBQ) \). The corresponding solution found is optimal by construction.

**2.6.1.2 A Single Level Optimization Model:**

Based on Proposition 4, the linear relaxation of followers’ problem provides an optimal solution. Thus, we replace followers’ problem with its linear relaxation. The corresponding formulation is presented
in Appendix A.3. We reformulate the followers’ problem by the corresponding KKT conditions to transform the bilevel optimization problem into a single level optimization model. The KKT conditions, which include the stationary constraints, primal feasibility constraints, dual feasibility constraints, and the complementary slackness constraints, are provided in Appendix A.3.

The corresponding single-level model is not linear due to the bilinear terms \( C_b X_{ibp} \) in the objective and similar terms in the constraints. Thus, we solved the single-level model using nonlinear solvers, such as Couenne (Belotti et al., 2009) and POD (Nagarajan et al., 2016, 2017). POD uses an adaptive, multivariate partitioning of bilinear terms. POD is an iterative algorithm which exploits the advantages of piecewise polyhedral relaxation approaches via disjunctive formulations to solve nonlinear programs to global optimality.

2.6.2 A Heuristic Solution Approach to Solve \((\bar{Q})\)

We propose an iterative procedure to solve \((\bar{Q})\). This procedure starts by initializing the prices set by the leader to \( \bar{C}_b = \min_{i \in \bar{I}_b, p=1} \{ t_{ib} + \bar{t}_{bp} \} \) for each \( b \in B \). Here, \( \bar{I}_b := I \). Given these prices, we solve followers’ problems \((\bar{Q}_{ib})\) to identify how much each follower should supply to maximize his profits. Let \( \bar{x}_f \) represent these quantities. At the initial step, only farm \( i^*_b \) (\( i^*_b = \arg\min_{i \in \bar{I}_b, p=1} \{ t_{ib} + \bar{t}_{bp} \} \)) would be offering biomass \( b \) to the biorefinery. Next, the leader solves her problem \((\bar{Q}_L)\) to identify whether the quantities provided by the followers satisfy her needs. If the leader’s problem is infeasible, then, she increases the prices offered. The prices are increased even when the leader’s problem is feasible in an effort to attract suppliers that provide products of higher quality. To find the new price, let \( \bar{I}_b = \bar{I}_b \setminus i^*_b \) and calculate \( \bar{C}_b = \min_{i \in \bar{I}_b, p=1} \{ t_{ib} + \bar{t}_{bp} \} \) for each \( b \in B \). If \( \bar{I} = \emptyset \), then, let \( p = p + 1 \), \( \bar{I}_b := I \) and calculate \( \bar{C}_b \). This procedure continues until no better solution is found for a fixed number of iterations \( \nu \).

\[
(\bar{Q}_L): \quad \min \bar{Z}^L = F(x, c^l) + \lambda \sum_{s=1}^N W_s + \mu \sum_{s=1}^N J_s \\
\text{s.t. Constraints} \quad (2.12) - (2.14) \\
c^l \geq 0, x \in \Pi_{i \in I} \lambda^F_i \\
x \leq \bar{x}^l.
\]

**Proposition 5.** Solutions found by the proposed heuristic are feasible for model \((\bar{Q})\).

It is straightforward to see that the proposed heuristic generates feasible solutions since the heuristic
stops when, at a given price offered by the leader, her problem is feasible; and the corresponding amounts of biomass offered maximize farmers’ profits.

2.6.3 A Lower Bound for \((\bar{Q})\)

Consider the following relaxation of model \((\bar{Q})\).

\[
(\bar{Q}^L): \quad \min \quad Z^L = F(x, c^l) + \lambda \sum_{s=1}^{N} W_s + \mu \sum_{s=1}^{N} J_s
\]

s.t. Constraints \((2.12)-(2.14)\)

\[c^l \geq 0, x \in \Pi_{i \in I} A_i^F.\]

This model minimizes the objective of the leader. The corresponding feasible region is the intersection of the feasible regions of the followers’ problems and the feasible region of the leader. Since this model formulation is a relaxation of \((\bar{Q})\), its optimal objective function value is a lower bound for \((\bar{Q})\). We use the corresponding lower bound in order to evaluate the performance of the heuristic solution approach proposed in Section 2.6.2.

2.7 Numerical Analysis

2.7.1 Case Study Description and Data

The main source of data is the Billion Ton Study (Langholtz et al., 2016). We focus our study in South Carolina, and use the county-level data of biomass supply. Table 2.1 lists the different types of biomass available in South Carolina which are suitable to use in a thermochemical conversion process. For each type of biomass and different scenarios, the report presents the expected amount available during 2014-2040. We use the data corresponding to these scenarios: “Medium housing, low energy demands” for forest biomass; “Wastes and other residues” for construction and demolition (C&D) and municipal solid waste (MSW); “Base case, single energy crops” for hybrid poplar. For poplar, we use the data of 2026, which is the most recent data available in the report. For the rest of biomass types, we use the data of 2016.

Tables 2.1 and 2.2 summarize the input data we use in the numerical analysis. This data is collected from (Jacobson et al., 2014a), Bioenergy Feedstock Library (INL, 2017) and (Harris et al., 2004). Table 2.1 presents the average ash content (AAC), the range of ash content before (ACR) and after (\(\overline{ACR}\)) pre-
processing, low heating value (LHV), and high heating value (HHV) for different types of biomass; and harvesting and collection (H&C) cost, processing cost (Pr) storage cost (St), fixed \((g_b)\) and variable \((v_b)\) transportation costs.

Table 2.1: Summary of Input Data

<table>
<thead>
<tr>
<th>Physical Properties</th>
<th>AAC (wt.%)</th>
<th>ACR (wt.%)</th>
<th>LHV (\times 10^6) BTU/DT</th>
<th>HHV (\times 10^6) BTU/DT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid Poplar</td>
<td>0.50</td>
<td>0.3 - 4.3</td>
<td>0.30 - 0.75</td>
<td>16.768</td>
</tr>
<tr>
<td>Pine</td>
<td>0.75</td>
<td>0.1 - 6.0</td>
<td>0.10 - 1.13</td>
<td>14.510</td>
</tr>
<tr>
<td>SP* Residue</td>
<td>1.00</td>
<td>0.8 - 2.2</td>
<td>0.80 - 1.50</td>
<td>15.232</td>
</tr>
<tr>
<td>SN** Residue</td>
<td>1.00</td>
<td>0.8 - 2.2</td>
<td>0.80 - 1.50</td>
<td>15.232</td>
</tr>
<tr>
<td>Mixed Residue</td>
<td>1.20</td>
<td>0.8 - 2.2</td>
<td>0.80 - 1.80</td>
<td>15.160</td>
</tr>
<tr>
<td>C&amp;D*** Waste</td>
<td>1.00</td>
<td>0.8 - 2.2</td>
<td>0.80 - 1.50</td>
<td>14.510</td>
</tr>
<tr>
<td>MSW**** Waste</td>
<td>10.00</td>
<td>7.0 - 15.0</td>
<td>7.00 - 15.00</td>
<td>10.250</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Costs</th>
<th>H&amp;C ($/DT)</th>
<th>Pr ($/DT)</th>
<th>St ($/DT)</th>
<th>(g_b) ($/DT)</th>
<th>(v_b) ($/DT/mile)</th>
<th>TQ (MDT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid Poplar</td>
<td>22.24</td>
<td>23.97</td>
<td>3.23</td>
<td>29.53</td>
<td>0.046</td>
<td>0.34</td>
</tr>
<tr>
<td>Pine</td>
<td>20.19</td>
<td>12.85</td>
<td>3.23</td>
<td>20.53</td>
<td>0.046</td>
<td>0.60</td>
</tr>
<tr>
<td>SP* Residue</td>
<td>0.00</td>
<td>23.97</td>
<td>3.23</td>
<td>20.69</td>
<td>0.046</td>
<td>0.11</td>
</tr>
<tr>
<td>SN** Residue</td>
<td>0.00</td>
<td>23.97</td>
<td>3.23</td>
<td>20.69</td>
<td>0.046</td>
<td>0.27</td>
</tr>
<tr>
<td>Mixed Residue</td>
<td>0.00</td>
<td>23.97</td>
<td>3.23</td>
<td>20.69</td>
<td>0.046</td>
<td>0.25</td>
</tr>
<tr>
<td>C&amp;D*** Waste</td>
<td>0.00</td>
<td>28.12</td>
<td>3.23</td>
<td>22.87</td>
<td>0.046</td>
<td>0.34</td>
</tr>
<tr>
<td>MSW**** Waste</td>
<td>0.00</td>
<td>19.70</td>
<td>4.50</td>
<td>20.69</td>
<td>0.046</td>
<td>0.099</td>
</tr>
</tbody>
</table>

* softwood planted, ** softwood natural, *** construction and demolition, **** municipal solid waste

In our numerical analysis we conduct sensitivity analysis with respect to biorefinery capacity. Table 2.2 presents biorefinery capacities in terms of thermal requirement and biomass supply.

Table 2.2: Biomass Refinery Requirements

<table>
<thead>
<tr>
<th>Demand (MDT/year)</th>
<th>Allowable Ash Content (%)</th>
<th>Thermal Requirement (\times 10^6) BTU/year</th>
<th>Thermal Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3 – 0.6</td>
<td>(\leq 1)</td>
<td>3,838 – 7,677</td>
<td>75 – 80</td>
</tr>
</tbody>
</table>

We consider each county to be a biomass supplier, and we assume that the available biomass is located at the centroid of this county. We consider that ash content for each supplier follows a triangular distribution with mean and range as defined in Table 2.1. We consider that biomass heating value and biomass type for each supplier is uniformly distributed with bounds defined by LHV and HHV presented in Table 2.1 (Shabani and Sowlati, 2016). Our model does not consider facility location decisions. Thus, we identify a biorefinery location in a separate model which minimizes the weighted travel distance to all suppliers.
The algorithms proposed are programmed in Julia 0.6.2 using modeling language JuMP (Dunning et al., 2017). These models ran on Clemson University’s high performance resource Palmetto Cluster and used 8 nodes and 64 GB RAM. The linear and mixed integer programs are solved using GUROBI callable subroutines.

2.7.2 Evaluating the Performance of the Solution Approaches

2.7.2.1 Linear Approximation Model ($\hat{P}$):

Table 2.3 presents the average, minimum, and maximum gap between the feasible solutions and the corresponding lower bounds of ($\bar{P}$) found by solving ($\hat{P}$) via the Algorithm for the Centralized Problem proposed in Section 2.4.3. The average (over 10 replications) error gap is less than 0.14% and the running time is smaller than 0.25 seconds which demonstrates that the proposed approximation can provide high quality solutions in a short amount of time.

<table>
<thead>
<tr>
<th>Demand</th>
<th>Avg</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.02</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>0.4</td>
<td>0.02</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>0.5</td>
<td>0.03</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>0.6</td>
<td>0.07</td>
<td>0.02</td>
<td>0.18</td>
</tr>
<tr>
<td>0.7</td>
<td>0.09</td>
<td>0.02</td>
<td>0.24</td>
</tr>
<tr>
<td>0.8</td>
<td>0.14</td>
<td>0.04</td>
<td>0.26</td>
</tr>
</tbody>
</table>

2.7.2.2 Single-Level Optimization Model to Solve ($\bar{Q}$):

The single-level optimization model is not linear due to a few nonlinear terms in the objective and constraints. In order to solve this Mixed-Integer Nonlinear Program (MINLP) we used Couenne and POD. Both solvers failed to solve instances with more than two suppliers, one biomass type and two cost brackets. Thus, we created a small size problem instance with two suppliers who supply hybrid poplar. We picked hybrid poplar since it has low ash content. For this problem we only generated two scenarios. The results are summarized in Table 2.4.

The results indicate that the running time of the proposed Heuristic (see Section 2.6.2) is order of magnitude smaller than the running time of Couenne and POD. The Heuristic also provides a feasible solution of higher quality which is 2% lower than the solution found from Couenne and 6% lower than the solution
Table 2.4: Evaluating the Single-Level Optimization Model for A Small Size Problem ($\beta = 0.1, \gamma = 0.1$)

<table>
<thead>
<tr>
<th>MINLP Solver</th>
<th>Demand Costs (MDT/year)</th>
<th>Costs ($/DT)</th>
<th>Ash Violation (%)</th>
<th>Thermal Running Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Couenne 0.3</td>
<td>0.3 128.59 0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>267.11</td>
</tr>
<tr>
<td>POD 0.3</td>
<td>0.3 133.48 0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>181.83</td>
</tr>
<tr>
<td>Heuristic</td>
<td>0.3 126.00 0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>6.45</td>
</tr>
</tbody>
</table>

found by POD. Notice that we use the “trunk” version of Couenne solver, thus the solution found is not guaranteed to be optimal (see (Belotti, 2018)), which is the case with the problem solved.

2.7.2.3 Heuristic Algorithm to Solve ($\hat{Q}$):

In order to evaluate the performance of the proposed Heuristic in solving large problem instances, we compare its solutions with the lower bounds found from solving ($\hat{Q}^L$) (see Section 2.6.3). ($\hat{Q}^L$) is a bilinear program which we solve using Couenne. Since the time it takes to solve this problem in Couenne is too long, we only solved the following problems. Problem 1 considers the whole dataset. Problem 2 considers a smaller dataset consisting only of suppliers of hybrid poplar and softwood residues. Problem 3 considers a supply chain with 23 suppliers (rather than the 46 suppliers we have in our dataset). We present the objective function value obtained from solving these problems and the corresponding error gap.

When solving Problem 1, we stopped Couenne after 48 hours. The solution found is not optimal. For Problems 2 and 3, we stopped Couenne after 10 hours. The solutions found for both problems are not optimal. Thus, we cannot claim that the objective function values of ($\hat{Q}^L$) are valid lower bounds for the Heuristic. However, the objective function values of Problems 1 and 2 are within 1% of the objective function value of the Heuristic.

Table 2.5: Comparison of ($\hat{Q}^L$) formulation and the Heuristic Algorithm ($N = 1$)

<table>
<thead>
<tr>
<th>Problem 1</th>
<th>Problem 2</th>
<th>Problem 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>($\hat{P}^L$)</td>
<td>Heuristic</td>
<td>Error Gap (in %)</td>
</tr>
<tr>
<td>Run Time (sec)</td>
<td>172,834</td>
<td>61</td>
</tr>
</tbody>
</table>
2.7.3 Managerial Insights

Tables 2.6 and 2.7 summarize the results of solving \((P)\) for different levels of demand.

Based on these results, SN residues contributes 30 to 45% of the amount in a blend and pine contributes 16 to 18%, and SP residues contribute 13 to 25% of the amount blended. MSW waste has not been utilized and the use of hybrid poplar increases with demand for biomass. Since hybrid poplar is expensive, this increase in utilization impacts the unit cost of the blend.

Comparing the results of Tables 2.6 and 2.7, one can observe that as the risk level decreases (i.e. \(\beta\) and \(\gamma\) decrease), the chance constraints become more restrictive, thus, the blends identified contain greater amounts pine and hybrid poplar since these biomass types have the lowest ash contents. The running time of the heuristic algorithm is higher when \(\beta = \gamma = 0.2\) (compared to \(\beta = \gamma = 0.3\)) since the sample size \(N\) is larger.

Table 2.6: Costs and Biomass Blending Ratios Under Ash and Thermal Content Uncertainties \((\beta = \gamma = 0.3)\)

<table>
<thead>
<tr>
<th>Demand (MDT/year)</th>
<th>Cost ($1000)</th>
<th>Costs ($/DT)</th>
<th>Blending Ratios for Thermochemical Conversion Process (in %)</th>
<th>Running Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hybrid Poplar</td>
<td>Pine</td>
</tr>
<tr>
<td>0.3</td>
<td>29,396</td>
<td>87.76</td>
<td>6.59</td>
<td>18.54</td>
</tr>
<tr>
<td>0.4</td>
<td>39,462</td>
<td>88.49</td>
<td>8.11</td>
<td>17.55</td>
</tr>
<tr>
<td>0.5</td>
<td>49,678</td>
<td>89.31</td>
<td>9.85</td>
<td>16.29</td>
</tr>
<tr>
<td>0.6</td>
<td>60,039</td>
<td>90.07</td>
<td>10.97</td>
<td>15.77</td>
</tr>
<tr>
<td>0.7</td>
<td>70,519</td>
<td>90.70</td>
<td>11.58</td>
<td>15.62</td>
</tr>
<tr>
<td>0.8</td>
<td>81,123</td>
<td>91.23</td>
<td>11.47</td>
<td>16.34</td>
</tr>
</tbody>
</table>

Note: \(\hat{\beta} = \hat{\gamma} = 0.3\) and \(N = 125\).

Table 2.7: Costs and Biomass Blending Ratios Under Ash and Thermal Content Uncertainties \((\beta = \gamma = 0.2)\)

<table>
<thead>
<tr>
<th>Demand (MDT/year)</th>
<th>Cost ($1000)</th>
<th>Costs ($/DT)</th>
<th>Blending Ratios for Thermochemical Conversion Process (in %)</th>
<th>Running Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hybrid Poplar</td>
<td>Pine</td>
</tr>
<tr>
<td>0.3</td>
<td>29,519</td>
<td>88.03</td>
<td>6.73</td>
<td>19.47</td>
</tr>
<tr>
<td>0.4</td>
<td>39,614</td>
<td>88.88</td>
<td>9.19</td>
<td>17.08</td>
</tr>
<tr>
<td>0.5</td>
<td>49,842</td>
<td>89.58</td>
<td>10.53</td>
<td>16.23</td>
</tr>
<tr>
<td>0.6</td>
<td>60,220</td>
<td>90.31</td>
<td>11.65</td>
<td>15.52</td>
</tr>
<tr>
<td>0.7</td>
<td>70,755</td>
<td>91.02</td>
<td>12.49</td>
<td>15.25</td>
</tr>
<tr>
<td>0.8</td>
<td>81,404</td>
<td>91.64</td>
<td>13.01</td>
<td>15.33</td>
</tr>
</tbody>
</table>

Note: \(\hat{\beta} = \hat{\gamma} = 0.2\), and \(N = 250\).

Tables 2.8 and 2.9 summarize the results from solving \((\bar{Q})\) for different levels of demand. Via these experiments we find that SN residues and pine count for about 50-60% of the blends identified. This is mainly because pine has low ash content and SN residues are not expensive. As demand increases, we observe an increase in the amount of pine used. This increase impacts the cost of the blend. The amount of mixed
residues and C&D waste also increase with demand. This is mainly due to the low cost of delivering these biomass types, and the limited amount of SN residues available. MSW was not used in a blend due to its high ash content. Hybrid poplar is used in moderation due to its high cost, although its ash content is low. Based on these results, decreasing risk level (i.e., $\beta, \gamma$ decrease) leads to higher costs in the supply chain.

Table 2.8: Costs and Biomass Blending Ratios Under Ash and Thermal Content Uncertainties ($\beta = \gamma = 0.3$)

<table>
<thead>
<tr>
<th>Demand (MDT/year)</th>
<th>Cost ($1000)</th>
<th>Cost ($/DT)</th>
<th>Blending Ratios for Thermochemical Conversion Process (in %)</th>
<th>Running Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hybrid Poplar Pine SP Residue SN Residue Mixed Residue C&amp;D Waste MSW</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>30,463</td>
<td>91.11</td>
<td>7.61 16.42 26.05 46.30 2.05 1.58 0.00</td>
<td>708</td>
</tr>
<tr>
<td>0.4</td>
<td>41,159</td>
<td>93.39</td>
<td>9.56 19.93 5.46 44.49 11.74 10.19 0.00</td>
<td>773</td>
</tr>
<tr>
<td>0.5</td>
<td>51,651</td>
<td>93.85</td>
<td>7.99 21.97 6.40 38.60 12.21 14.51 0.00</td>
<td>750</td>
</tr>
<tr>
<td>0.6</td>
<td>62,403</td>
<td>94.46</td>
<td>7.26 22.98 7.44 34.05 12.94 17.08 0.00</td>
<td>777</td>
</tr>
<tr>
<td>0.7</td>
<td>73,305</td>
<td>95.12</td>
<td>7.55 22.47 8.41 30.52 13.21 19.58 0.00</td>
<td>703</td>
</tr>
<tr>
<td>0.8</td>
<td>84,175</td>
<td>95.58</td>
<td>7.17 23.32 8.71 27.56 13.96 21.14 0.00</td>
<td>696</td>
</tr>
</tbody>
</table>

Note, $\hat{\beta} = \hat{\gamma} = 0.0, N = 125$.

Table 2.9: Costs and Biomass Blending Ratios Under Ash and Thermal Content Uncertainties ($\beta = \gamma = 0.2$)

<table>
<thead>
<tr>
<th>Demand (MDT/year)</th>
<th>Cost ($1000)</th>
<th>Cost ($/DT)</th>
<th>Blending Ratios for Thermalchemical Conversion Process (in %)</th>
<th>Running Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hybrid Poplar Pine SP Residue SN Residue Mixed Residue C&amp;D Waste MSW</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>30,585</td>
<td>91.54</td>
<td>8.93 15.82 26.07 46.18 1.19 1.80 0.00</td>
<td>1,940</td>
</tr>
<tr>
<td>0.4</td>
<td>41,096</td>
<td>93.74</td>
<td>9.78 20.91 6.27 44.91 10.27 9.89 0.00</td>
<td>1,947</td>
</tr>
<tr>
<td>0.5</td>
<td>51,664</td>
<td>94.33</td>
<td>8.89 22.16 7.25 39.35 11.87 12.68 0.00</td>
<td>1,920</td>
</tr>
<tr>
<td>0.6</td>
<td>62,543</td>
<td>95.03</td>
<td>9.05 21.70 8.42 35.12 11.96 15.78 0.00</td>
<td>1,789</td>
</tr>
<tr>
<td>0.7</td>
<td>73,641</td>
<td>95.67</td>
<td>9.11 21.83 8.96 31.45 13.19 17.22 0.00</td>
<td>1,729</td>
</tr>
<tr>
<td>0.8</td>
<td>84,841</td>
<td>95.98</td>
<td>7.17 24.79 9.43 27.98 13.25 18.99 0.00</td>
<td>1,646</td>
</tr>
</tbody>
</table>

Note, $\hat{\beta} = \hat{\gamma} = 0.0, N = 250$.

Figure 2.3 summarizes the gap between the objective function value of the centralized and decentralized models. The results indicate that the centralized model ($P$) provides solutions which have lower costs compared to the decentralized model ($Q$). The gap presented varies between 2 and 6%. This difference is due to supplier selection and biomass purchasing (blending) decisions. Based on the results of the sensitivity analysis, both models suggest that biomass blends should consist of 30 to 45% SN residues and 0% MSW. However, the suggested percentages for the rest of biomass types differ. This is because the objectives of both models are different. The centralized model minimizes system-wide costs, and the decentralized system minimizes biorefinery costs. This later objective leads to higher supply chain costs.

These results point to the estimation errors when assuming centralized decision making. The decentralized model, which is more realistic, leads to higher supply chain costs.
Figure 2.3: Decentralize vs Centralize - Percent Gap

(a) \( \beta = 0.2, \gamma = 0.2 \)

(b) \( \beta = 0.3, \gamma = 0.3 \)
Chapter 3

Optimization Models for Integrated Biorefinery Operations

3.1 Background

In the last two decades, significant investments have been made by a number of federal agencies to establish and develop the cellulosic biofuel industry in the U.S. Despite of these investments, this industry remains a nascent concern. Major challenges faced by this industry are variations of biomass supply and variations of physical/chemical characteristics of biomass (e.g., ash, moisture, carbohydrate contents, etc.), largely due to spatial variations in weather and soil, harvesting equipment used, etc. These variations lead to uneven flow of biomass in a biorefinery which affect equipment utilization rate and lead to inconsistent conversion rates. Variations in biomass supply and processing costs increase the risk of investing in this industry.

A recent report from the US Department of Energy (DOE) identifies “bulk solids handling and material flows through the system” as a critical component to achieve the design throughput of the conversion processes (U.S. Department of Energy (DOE), 2016). The flow of materials is impacted by variations of biomass characteristics. For example, consider the scenario when a number of bales of different types of feedstock with different moisture level and ash content are processed on the same equipment (i.e., grinder). The resulting distribution of particle size and particle uniformity of processed biomass varies from one bale to the next. These variations negatively affect the integration of biomass feeding system and conversion process,
which leads to low/unreliable on-stream time of equipment and low utilization of the reactor.

The main objective of this chapter is to improve the reliability of biomass feeding system in a biorefinery via a stochastic optimization model. The reliability of the biomass feeding system is defined as the probability of achieving the targeted reactor utilization rate under stochastic biomass characteristics and stochastic equipment failure. In this study, we only consider equipment failures caused by clogging, overflow (in conveyors), and overheat. These failures are mainly due to biomass characteristics, such as, moisture content and particle size distribution. We focus on the reactor because it is the most expensive equipment in a biorefinery. Achieving the targeted utilization rate at all time is expensive and not practical, because it requires the use of biomass with low moisture and ash contents. Therefore, it is economical to focus on achieving the utilization of reactor at a target rate, most of the time. We model this requirement using chance constraints.

In order to optimize the performance of the reactor, the proposed optimization model identifies optimal equipment operating conditions, inventory level for a given sequence of bales based on biomass moisture levels. These approaches align with strategies presented in the literature, which indicate that, system’s reliability can be improved by increasing redundancy, sequencing components and products, etc. (Coit and Zio, 2018). In our system, bale sequencing impacts the size of inventory, and inventory holding mitigates disruption of biomass flow due to failure and operating conditions of equipment.

Biomass density is one of the main factors impacting the biomass flow in the system. Studies show that biomass particle size distribution and moisture content significantly affect biomass density (Zhou et al., 2008), (Crawford et al., 2016). To evaluate these relations we develop a Discrete Element Method (DEM) model. DEM is a computational model which describes the mechanical bulk behaviour of granular materials (Cundall and Strack, 1979). DEM simulates the movement and interaction of particles with each other and with the system. The results of this simulation are used to develop regression functions which calculate the density of biomass as a function of moisture level and particle size distribution. We use biomass density to calculate biomass flow in the system. The proposed optimization uses a multi-period network flow model to capture the flow of biomass in the system, and the flow of biomass into the reactor during the planning horizon.

We develop a case study using historical data from the Process Development Unit (PDU), a biomass facility, at Idaho National Laboratory (INL). We use this case study to validate the proposed model and conduct numerical experiments. In particular, our numerical experiments are designed to evaluate the impact of (i) biomass characteristics (e.g., moisture level, particle size distribution and density) on achieving the
targeted reactor utilization rate and minimizing operating costs; (ii) equipment failure on reactor utilization rate and costs; and (iii) strategies, such as, sequencing bales, changing operating condition of equipment, and keeping inventory, on reducing the risk of achieving the targeted reactor utilization rate.

We expect that the results of this study will help biorefineries to develop strategies which lead to increased reactor utilization and minimize costs; and to identify optimal operational condition in face of stochastic biomass characteristics and equipment failure. These outcomes will facilitate commercial scale generation of biofuels at competitive cost. In the long run, these outcomes will strengthen the sustainable bioeconomy of US, enhance the security of energy supplies, reduce dependencies on fossil fuels, and reduce greenhouse gas (GHG) emissions. A strong sustainable bioeconomy has additional socioeconomic benefits, such as generating new green jobs, growth of rural economy and social stability, among others (Domac et al., 2005; Sims, 2003; You et al., 2012).

The remainder of the chapter is organized as follows. In Section 3.2, we review the literature. In Section 3.3, we describe the problem and the modeling approach. A case study is presented in Section 3.4. In Section 3.5, we present numerical results and analysis.

### 3.2 Literature Review

The two main streams of literature closely related to this study are system reliability and optimization of biorefinery operations. There are a number of studies in the literature that focus on biomass supply chain optimization (Chen and Önal, 2014; Ekşioğlu et al., 2009; Memişoğlu and Üster, 2016; Roni et al., 2014). Supply chain decisions impact the availability of biomass in a biorefinery. In this work we do not consider supply chain decisions and assume that the mix and quantity of biomass to process is given as an input to the problem.

#### 3.2.1 System Reliability

Maximizing system reliability is a relevant problem in manufacturing and service systems that operate under uncertainty. During the last decade, a number of researchers studied the problem and developed solution approaches. The field of system reliability optimization is still expanding. A recent survey paper by (Coit and Zio, 2018) presents a comprehensive review of this literature. It provides a chronological grouping of the research in the field into three eras, the era of mathematical programming, era of pragmatism, and era of active reliability improvement. It also classifies the problems in this field into four major groups (i)
redundancy allocation, (ii) reliability allocation, (iii) reliability-redundancy allocation, and (iv) assignment and sequencing. Each group of research provides strategies to improve the system’s reliability. The model we propose contributes to redundancy allocation and sequencing streams of research since it identifies the amount of inventory of biomass and the sequencing of bales processed which minimize the cost of meeting a targeted system reliability level.

A wide range of methods are used to solve reliability problems. For example, (Bellman, 1966; Fyffe et al., 1968; Ghare and Taylor, 1969) use dynamic programming to solve a redundancy allocation problem. Work by (Misra, 1991; Prasad and Kuo, 2000) use linear programming and integer programming models to minimize costs or maximize reliability. Many system reliability problems are combinatorial in nature due to component selection and sequencing. Additionally, models of complex systems use nonlinear constraints and objectives. Therefore, exact solution approaches are computationally expensive. To address these computational challenges, a number of metaheuristics are proposed, such as, genetic algorithms and simulated annealing, which are shown to be efficient and robust. For example, (Painton and Campbell, 1995) developed a genetic algorithm to model a personal computer component configuration problem, (Yang et al., 1999) use a genetic algorithm to design a nuclear power plant. While metaheuristics do not guarantee to find the global optima, researchers showed that metaheuristics are efficient and robust methods to solve complex system reliability problems.

Recent developments capture the impact of uncertainty to the reliability of a system. For example, (Basciftci et al., 2018; Li et al., 2008) use chance-constrained stochastic programming to model and solve a system reliability problem under uncertainty. Similar to this work, our proposed model uses chance constraints to capture the impact of biomass characteristics and equipment failure on the utilization of the reactor.

Current efforts in this field are focused on the development of data driven optimization models (Yang et al., 2010; Alsina et al., 2018; Meng et al., 2020). The proposed work also contributes to this stream of research by using sensor-based data collected from the PDU to develop the DEM models and estimate failure probabilities of equipment.

### 3.2.2 Optimization of Biorefinery Operations

Biorefinery operations include biomass pre-processing, biomass handling and biomass storage. Most of the literature related to biorefinery operations focuses on the evaluation of design parameters of the equipment used (Crawford et al., 2016; Dai et al., 2012). Other studies analyze energy consumption of an equipment as a function of its design parameters and biomass characteristics (Jacobson et al., 2014b; Kenney
et al., 2014; Yancey and Tumuluru, 2015). These studies are limited in scope since they do not capture the interactions among equipment and the impact of equipment on the performance of the system as a whole.

A number of studies use optimal control models to minimize the energy consumption of an equipment. For example, (Numbi and Xia, 2016) propose a deterministic, non-linear optimization model that identifies optimal operating parameters of a crusher. By introducing a time-to-use electricity tariff, the model achieves additional cost savings. Work by (Zhang and Xia, 2010) presents an optimal control model for a series of conveyor belts. The authors compare two different operational structures, one in which the conveyors’ speed is fixed, and another in which the conveyors’ speed is adjusted over time. The comparison of these models points to the benefits of having the flexibility to update the operating conditions of equipment. To summarize, optimal control models are complex non-linear programs, thus, are used to model only parts of a system. Using such an approach to model the performance of a system under uncertainty, would lead to a notoriously complex model. This is the main reason why we did not follow such an approach.

A few studies present models for an optimal design of operations in a biorefinery. For example, (Pham and El-Halwagi, 2012) propose a two steps approach to optimize the design and operations of a biorefinery. This model identifies an optimal biorefinery configuration for a given set of biomass feedstocks and available conversion technologies. The first step, called the bi-directional synthesis, identifies the intermediary chemicals that can be produced using the feedstocks available. The second step optimizes a network flow problem that identifies the mix of intermediary chemicals and corresponding conversion technologies that minimize the total production cost. Work by (Zondervan et al., 2011) uses a mixed-integer and nonlinear network optimization model to identify an optimal design for a biorefinery. The model identifies process sequences to optimize production of a set of biofuels and bioproducts. Both works, (Pham and El-Halwagi, 2012) and (Zondervan et al., 2011), use deterministic models since they consider fixed biomass characteristics, flow rates and yields. Different from these works, our research models uncertainties in this system. The use of our proposed models leads to robust decisions.

A number of studies use DEM models to study the flowability of materials. DEM simulates movements and interactions of particles with each other and with the system. By using the law of force, DEM calculates the rotation and velocity of each particle and simulates the movement of material in an equipment (or system) realistically (Höchner et al., 2012). Additionally, DEM models consider the impact of bond strengths of particles and deformability of materials to flowability (Xia et al., 2019). The main reason for the growing interest in these models is the lack and inaccuracy of experimental data necessary to model the flowability of granular material (Xia et al., 2019). In the last decade DEM models have been used to evaluate the flowability
of different biomass feedstocks (Xia et al., 2019; Guo et al., 2020). A study by (Scherer et al., 2016) develops a DEM model to simulate the drying process of wood chips in a rotary dryer. They analyzed the impacts of equipment design on the process rate and quality. (Orefice and Khinast, 2017) develops a DEM model of a horizontal screw conveyor, a common transportation equipment for biorefineries. They study the impacts of screw rotational velocity and initial filling level of the conveyor on the flow rate of the material. (Xia et al., 2019) developed a DEM model to study the impacts of deformable pinewood chips in cyclic loading test. This study (Xia et al., 2019) shows that DEM can be used to simulate particle deformation, which is critical to the performance of equipment used in a biorefinery (e.g., grinders and pelleting equipment). Work by (Guo et al., 2020) present a bonded-sphere DEM model designed for switchgrass particles. This model generates data which are used to evaluate the relationship between biomass density and moisture level and particle size distribution. Our proposed work uses the results of this model to estimate material density. This integration of the results of the DEM model within the optimization model is one of the contributions of the proposed work.

### 3.3 Problem Description and Formulation

The goal of the proposed mathematical model is to identify process controls and inventory levels which minimize system-wide costs while meeting the desired biomass specifications for biochemical conversion; and achieving the designed production throughput. Our model is developed using data from the PDU since it provides a good representation of the processes used for feeding of biomass to the reactor in a biorefinery. Next, we describe the problem, present a mathematical formulation, present a solution approach and model extensions.

#### 3.3.1 Problem Description

Figure 3.1 presents the flowchart of the different processes of PDU. In this figure, notations $x_{1t}, x_{2t}, \ldots$ represent the flow of biomass in period $t$. Bales of biomass enter the system, one at a time, via a conveyor belt. Biomass is processed in Grinder 1. Processes biomass undergoes a process which separates biomass based on particle size. Large biomass particles are transported via conveyors to Grinder 2 for further processing. Small biomass particles are transported to the Metering bin for storage. Biomass processed in Grinder 2 also is stored in the Metering bin. From the Metering bin biomass is transported to the pelleting machine. Pelleted biomass is fed to the reactor.
We group the equipment in this system based on their tasks into processing, transportation and storage. A processing equipment converts biomass from its original format (e.g. baled, log, or coarse-shredded) to its final format (e.g., ground biomass and pellet). These equipment are grinders 1 and 2 and the pelleting mill. Most equipment of PDU are transportation equipment, such as, conveyors. Storage equipment, such as, the metering bin, store in-process inventory.

The following are the assumptions we make, which facilitate the modeling of this system. First, historical data from PDU reports only the average moisture level of bales processed. This is calculated using data collected via sensors located in the conveyor belt which feeds biomass bales to the system. It takes a few minutes to process one bale. Since the moisture level of a bale is heterogeneous, we generate the moisture level of biomass in different time periods using a uniform distribution. The mean of this distribution equals the average moisture level of the corresponding bale. The lower and upper bounds of the distribution are established via discussions with experts.

Second, the moisture level of biomass is reduced after being processed inside a piece of equipment. Historical data is used to estimate these reductions in moisture level for each processing equipment.

Third, work by (Yancey and Tumuluru, 2015) shows that the probability of clogging increases with moisture level of biomass being processed. Thus, in practice, the processing speed of equipment is adjusted based on the moisture level of biomass. This impacts the processing capacity of equipment. Additionally, moisture level impacts density, which impacts the volume of biomass stored. Thus, the volumetric storage capacity of the metering bin depends on the moisture level of biomass stored. For these reasons, in our model...
the processing capacity of equipment is a function of the moisture level of biomass.

3.3.2 Problem Formulation

We use stochastic optimization to model the uncertainties observed in the system. This model is a two-stage stochastic program with chance constraints. We use a two-stage stochastic program because some of the decisions are made before uncertainties (such as, biomass moisture content, particle size, and bulk density) reveal. These first-stage decisions include the processing speed of equipment over the entire planning horizon, $V$, and the initial inventory level in the metering bin, $I_0$. Next, given a realization of the biomass characteristics $\omega$ and the first-stage decisions, the second-stage decisions are made, which include biomass flows in the system, $X(\omega)$, and the inventory level in the metering bin, $I(\omega)$. Note that, for ease of notation and to simplify our model formulation, we use vectors $X(\omega)$, $I(\omega)$, etc. in formulations presented in this Section. Explicit definitions of these vectors are presented in Table 3.1. Detailed notation and model formulation are presented in Appendices A and B.

We use a chance constraint to model the reliability level of the reactor. Since maintaining the reactor operating at the targeted utilization rate all the time is expensive, we aim to achieve this target utilization rate most of the time. Let $r$ represent this target rate and $1 - \epsilon$ the desired reliability of the reactor. We use a chance constraint to ensure that the reactor achieves the target rate $r$ at least $(1 - \epsilon) \times 100$ percent of the time.

We use network flow constraints to model the flow of biomass within the system in every period during the planning horizon $T$. Let $G = (N, A)$, with a node set $N$ and arc set $A$, represent the network structure of the system. The set of nodes represent equipment of the PDU (Figure 3.1), and the set of arcs represent the flow of biomass from one equipment to next. This network is rather simple since it represents processes of the PDU. Nevertheless, the network structure would be complex if used to model operations of medium and large size plants. The following is a list of notation we use.

For simplicity of presentation, we next present a succinct formulation $(P)$ using generic functional notation such as $f(\cdot)$, $b(\cdot)$ and $d(\cdot)$ to model the relationship between decision variables and random variables. A detailed formulation is provided in Appendix B. Given a sequence of bales with moisture level $\kappa$ to process in the planning horizon of $T$, the proposed stochastic program is given by:
Table 3.1: Mathematical Notation

<table>
<thead>
<tr>
<th>SETS:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>The set of moisture levels of biomass.</td>
</tr>
<tr>
<td>T</td>
<td>The set of time periods in the planning horizon, $T := {1, 2, \ldots, T}$</td>
</tr>
<tr>
<td>$E^p$</td>
<td>The set of processing equipment.</td>
</tr>
<tr>
<td>$E^r$</td>
<td>The set of transportation equipment.</td>
</tr>
<tr>
<td>$E^m$</td>
<td>The set of storage equipment.</td>
</tr>
<tr>
<td>$N = E^p \cup E^r \cup E^m$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARAMETERS:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>Node arc incidence matrix of $G$.</td>
</tr>
<tr>
<td>$A'$</td>
<td>Node arc incidence matrix of $G'(E^r, A')$.</td>
</tr>
<tr>
<td>$h$</td>
<td>Inventory holding cost for storage equipment (in $/ton).</td>
</tr>
<tr>
<td>$f(\cdot)$</td>
<td>Energy consumption and operational cost function.</td>
</tr>
<tr>
<td>$\kappa := {\kappa_t}_{t \in T}$</td>
<td>Moisture level of the biomass bales ($\kappa_t \in M$)</td>
</tr>
<tr>
<td>$\bar{v}(\kappa_i)$</td>
<td>Upper bound of equipment processing speed.</td>
</tr>
<tr>
<td>$\bar{i}(\kappa_0)$</td>
<td>Initial inventory holding capacity in the storage equipment.</td>
</tr>
<tr>
<td>$r$</td>
<td>The target reactor feeding rate (in tons).</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Risk tolerance parameter.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RANDOM PARAMETERS:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{i}(\omega)$</td>
<td>The inventory holding capacity in the storage equipment for processing biomass $\omega$.</td>
</tr>
<tr>
<td>$\underline{i}(\omega)$</td>
<td>The inventory lower threshold in the storage equipment for processing biomass $\omega$.</td>
</tr>
<tr>
<td>$\omega$</td>
<td>A random vector that contains the stochastic biomass characteristics.</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>The support of the probability distribution of $\omega$.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DECISION VARIABLES:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$V = {V_t}_{t \in T}$</td>
<td>Equipment processing speed (in meters or rotations per time period).</td>
</tr>
<tr>
<td>where $V_t := {V_{it}}_{i \in N}$, $\forall t \in T$</td>
<td></td>
</tr>
<tr>
<td>$I_0 := {I_0}_{i \in E^m}$</td>
<td>Initial inventory level in the storage equipment</td>
</tr>
<tr>
<td>$I(\omega) := {I_t(\omega)}_{t \in T}$</td>
<td>Inventory level when processing biomass with characteristics $\omega$ (in tons).</td>
</tr>
<tr>
<td>$X(\omega) := {X_t(\omega)}_{t \in T}$</td>
<td>Biomass flow from equipment when processing biomass with characteristics $\omega$ (in tons).</td>
</tr>
<tr>
<td>$X_t(\omega) := {X_{it}(\omega)}_{i \in N}$</td>
<td>Biomass flow from equipment $i$ when processing biomass with characteristics $\omega$ at period $t$ (in tons).</td>
</tr>
</tbody>
</table>

\[ (P) \quad \min : Z = h^T I_0 + \mathbb{E}[f(X(\omega), I(\omega), \omega)] \]  
\[ \text{s.t. } A' X(\omega) = 0, \ \forall \omega \in \Omega, \]  
\[ 0 \leq V_t \leq \bar{v}(\kappa_i), \ \forall t \in T, \]  
\[ X_t(\omega) = b(V_t, I_0, I(\omega), \omega), \ \forall t \in T, \omega \in \Omega, \]  
\[ X_t(\omega) \leq d(V_t, \omega), \ \forall t \in T, \omega \in \Omega, \]  
\[ \underline{i}(\omega) \leq I_t(\omega) \leq \bar{i}(\omega), \ \forall t \in T, \omega \in \Omega, \]  
\[ 0 \leq I_0 \leq \bar{i}(\kappa_0), \]  
\[ \mathbb{P}(R(X(\omega), \omega) \geq r) \geq 1 - \epsilon. \]  

The objective function (3.1) is the total expected cost of the system, which includes the initial inventory holding cost and the expected operating costs. The operating costs, $f(\cdot)$, include the cost of energy consumed by equipment during the planning horizon.

Constraints (3.2) represent the flow balance constraints of transportation equipment. Constraints
set an upper bound $\bar{v}(\kappa_t)$ on the processing speed of each equipment. These bounds depend on the moisture level $\kappa_t$ of biomass being processed at period $t$.

Constraints (3.4) calculate the flow from storage and processing equipment. For storage equipment, these represent the inventory balance constraints. Pelleting mill, different from other processing equipment, has a positive residence time. Therefore, an inventory balance constraint is used to calculate the flow of biomass from this equipment. The inventory balance constraint is written explicitly as follows:

$$I_{i,t-1}(\omega) + \sum_{j \in \delta_i^-} X_{jt}(\omega) = I_{it}(\omega) + X_{it}(\omega), \forall i \in E^m, t \in \mathcal{T} \setminus \{1\},$$

$$I_{i,0} + \sum_{j \in \delta_i^-} X_{jt}(\omega) = I_{i1}(\omega) + X_{i1}(\omega), \forall i \in E^m,$$

where $\delta_i^-$ represents the set of equipment that feeds into equipment $i$.

Constraints (3.4) are simpler for other processing equipment, such as grinders. Let $\tilde{d}_{it}(\omega)$ represent the density of biomass after processed in grinder $i$, and let $\gamma_i$ represent the cross section area of the discharge opening of grinder $i$, thus $\gamma_i \tilde{d}_{it}(\omega)$ is the mass discharging rate of the grinder. Given the infeed rate of the grinder $V_{it}$, the flow from a grinder is calculated as:

$$X_{it}(\omega) = \gamma_i \tilde{d}_{it}(\omega)V_{it}.$$

Constraints (3.5) represent the upper limit on the amount of biomass flow from equipment $i \in \mathcal{N}$. For example, let $i$ be a screw conveyor with cross section area equal to $\gamma_i$. Thus, $\gamma_i \tilde{d}_{it}(\omega)$ is the mass discharging rate, and $\gamma_i \tilde{d}_{it}(\omega)V_{it}$ is the maximum amount of biomass that can be discharged by the conveyor.

$$X_{it}(\omega) \leq \gamma_i \tilde{d}_{it}(\omega)V_{it}.$$

Constraints (3.6) and (3.7) set upper and lower bounds in the amount of biomass that is stored in the metering bin. These bounds are random in (3.6) because the mass of biomass that can be stored depends on its density, which is a random parameter. We set lower bounds since this is a practice used at PDU to maintain consistent flow of biomass from the metering bin. Finally, (3.8) is the chance constraint.
3.3.3 A Sample Average Approximation of (P)

Chance-constrained stochastic programs are typically difficult to solve. This is because, for a given solution, it is often difficult to check its feasibility since it may require multi-dimensional integration. Furthermore, the feasible region defined by chance constraints is in general non-convex. To address these challenges, many researchers use the Sample Average Approximation (SAA) method to approximate these constraints. The corresponding model is an integer program which is easier to solve using commercial solvers (Atlason et al., 2008; Luedtke and Ahmed, 2008; Pagnoncelli et al., 2009).

The SAA model replaces the true probability distribution of random parameters with an empirical distribution obtained from random samples. Let \( \{\omega_1, \omega_2, \ldots, \omega_S\} \) be a set of \( S \) independent and identically distributed realizations (scenario) of \( \omega \), which are obtained via a Monte Carlo simulation. Thus, the probability associated with each of these realizations equals \( \frac{1}{S} \). In addition, let the second-stage decision variables \( X_t(\omega_s) \) and \( I_t(\omega_s) \) be denoted by short-handed notation \( X_{ts} \) and \( I_{ts} \), respectively, representing the flow value and inventory level in each period \( t \) under each scenario \( s \). Let \( \mathbb{1}() \) be an indicator function. The SAA approximation of \( (P) \) can be written as:

\[
\begin{align*}
(\hat{P}) \quad Z^S &= \min \ h^T I_0 + \frac{1}{S} \sum_{s=1}^{S} f(X_s, I_s, \omega_s) \\
\text{s.t.} \quad & (3.3), (3.7), \quad (3.9) \\
& A'X_s = 0, \ s = 1, 2, \ldots, S, \quad (3.10) \\
& X_{ts} = b(V_t, I_0, I_{ts}, \omega_s), \ \forall t \in T, s = 1, 2, \ldots, S, \quad (3.11) \\
& X_{ts} \leq d(V_t, \omega_s), \ \forall t \in T, s = 1, 2, \ldots, S, \quad (3.12) \\
& \mathbb{1}(\omega_s) \leq I_{ts} \leq \bar{\mathbb{1}}(\omega_s), \ \forall t \in T, s = 1, 2, \ldots, S, \quad (3.13) \\
& \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}[R(X_s, \omega_s) \geq r] \geq 1 - \hat{\epsilon}.
\end{align*}
\]

In formulation \( (\hat{P}) \), the value of \( \hat{\epsilon} \) may be different from the true risk parameter \( \epsilon \) used in formulation \( (P) \). Based on (Luedtke and Ahmed, 2008), when \( \hat{\epsilon} < \epsilon \), the probability that a feasible solution of \( (\hat{P}) \) is feasible to \( (P) \) increases with the sample size \( S \). In our experiment we choose to use \( \hat{\epsilon} < \epsilon \) in experiments where we consider equipment failures and use \( \hat{\epsilon} = \epsilon \) elsewhere.

Constraint (3.13) of \( (\hat{P}) \) is not linear because it includes an indicator function. The following is an equivalent mixed integer programming (MIP) reformulation of \( (\hat{P}) \) where constraint (3.13) is replaced by
and (3.16):

\[
\bar{Z}^S = \min \ h^\top I_0 + \frac{1}{S} \sum_{s=1}^{S} f(X_s, I_s, \omega_s)
\]

s.t. (3.3), (3.7), (3.9) – (3.12),

\[
R(X_s, \omega_s) \geq rz^s, \ s = 1, 2, \ldots, S,
\]

\[
\sum_{s=1}^{S} z^s \geq S(1 - \epsilon),
\]

\[
z^s \in \{0, 1\}, \ s = 1, 2, \ldots, S.
\]

However, solving this MIP it can be computationally expensive (Luedtke and Ahmed, 2008). Instead, we use a heuristic approach (Charnes et al., 1955; Abdelaziz, 2012) to solve (\( \hat{P} \)) which penalizes the violation of constraint (3.13). To this end, we introduce variables \( U \) to quantify the amount of violations, and a penalty parameter \( \pi > 0 \). The following is a linear approximation of (\( \hat{P} \)).

\[
\bar{Z}^S = \min \ h^\top I_0 + \frac{1}{S} \sum_{s=1}^{S} f(X_s, I_s, \omega_s) + \sum_{s=1}^{S} \pi U_s
\]

s.t. (3.3), (3.7), (3.9) – (3.12),

\[
R(X_s, \omega_s) + U_s - J_s = r, \ s = 1, 2, \ldots, S,
\]

\[
U_s, J_s \geq 0, \ s = 1, 2, \ldots, S.
\]

The new term in the objective function of (\( \bar{P} \)) penalizes the difference between the average amount of biomass fed to the reactor and the target value \( r \), if the target cannot be achieved.

For a given value of the penalty parameter \( \pi > 0 \), formulation (\( \bar{P} \)), which is a two-stage stochastic linear program, is easy to solve. Note that, parameter \( \pi \) is not known in advance and one needs to identify an appropriate \( \pi \) value so that the resulting solution satisfies the chance constraint \( (1 - \hat{\epsilon}) \times 100 \) of the time. It is obvious that \( \pi \to \infty \) leads to solutions for which \( U_s = 0 \). In this case, the chance constraint is satisfied for all \( s \). Conversely, \( \pi \to 0 \) leads to solutions for which \( U_s \geq 0 \). In this case, the chance constraint may be satisfied in less then \([1 - \hat{\epsilon}]S\) of the scenarios generated. Thus, we design a bisection search algorithm which identifies the smallest value of \( \pi \) for which the number of scenarios which satisfy the chance constraint (i.e. \( R(X_s, \omega_s) \geq r \)) is close to \([1 - \hat{\epsilon}]S\). We present this algorithm in detail in the Appendix D.
We use the stochastic Benders decomposition approach to solve the corresponding two-stage stochastic linear program ($\tilde{P}$) (Birge and Louveaux, 2011). We use Benders decomposition because the problem size increases with the number of scenarios. We use the multi-cut version of Benders decomposition.

### 3.3.4 Model Extensions

**Integration of the DEM Models.** Density impacts the flow of biomass in the system. Studies show that biomass particle size distribution and moisture content significantly affect its density (Crawford et al., 2016; Zhou et al., 2008). Using the simulation results of a DEM model, a regression function is developed that captures the relationships among biomass particle size distribution, moisture content, and biomass density accurately and realistically. Notice that, DEM models are computationally expensive to simulate. Thus, often, the results of the DEM model do not cover the full range of observations. By developing a regression function, we were able to extrapolate the DEM results to all possible scenarios considered in problem $\tilde{P}$.

One of the DEM models developed by (Guo et al., 2020) simulates grinders 1 and 2. This model is calibrated and validated using historical data from the PDU. The model estimates the bulk density of biomass via simulation. The simulation results are used to develop regression functions that capture the relationship between biomass density (dependent variable), and moisture level and particle size distribution (independent variables). Let $\rho_{j}^{i}$ represent the $j$-th percentile of the particle size distribution of biomass processed in equipment $i$. The regression model is given by:

$$\tilde{d}_{it} = \alpha_0^i + \alpha_1^i m_{it} + \alpha_2^i \rho_{50}^i + \alpha_3^i \rho_{10}^i + \tilde{\xi}_i,$$

where $\tilde{\xi}_i$ corresponds to a random error term following a normal distribution with mean equal to zero and a constant standard deviation. This regression model is used to compute the density of biomass in the proposed model for each of the scenarios generated.

**Incorporating equipment failure.** The proposed model can also be extended by incorporating random equipment failures. We consider two types of equipment failures: short-duration and long-duration. Short-duration failures are due to the overfeeding of the system, overflowing of conveyors, or overheating of the grinders. Long-duration failures are due to the clogging of equipment, which typically happens when processing biomass with high moisture level and large particle size. We assume that short-duration and long-duration failures are independent of each other.

For a short-duration failure, we assume that the system restarts automatically within a few seconds.
A long-duration failure does typically require an operator to unclog the equipment and restart the system. These failures are observed in equipment located in the upstream of the metering bin (the storage equipment), such as the grinders and the corresponding conveyors. Failures are not observed in the downstream of the metering bin since the flow of biomass is controlled via the inventory. Additionally, the moisture level of biomass is reduced considerably by the time it leaves the metering bin, and particle size of biomass is reduced considerably after being processed in grinder 2. Both factors reduce the probability of clogging.

To model the random equipment failures, we assume that the time between consecutive equipment failures and time to repair are random variables following certain probability distributions. We assume that the time between failures follows a Weibull distribution, and the duration of a failure follows a uniform distribution. The parameters used for these distributions are summarized in Table 3.2, which are validated using data collected from the PDU via maximum likelihood estimation method. We generate an offline operational schedule of equipment (with up and down time) over the planning horizon using random samples according to these distributions. These random operating schedules are incorporated in the model via parameters $o_{it}(\omega) \in \{0, 1\}$, where, $o_{it}(\omega) = 0$ if equipment $i$ is down in period $t$, and $o_{it}(\omega) = 1$ if the equipment is operating. We update constraints (3.4) which calculate the flow of biomass from equipment as follows:

$$X_{it}(\omega) = o_{it}(\omega)\gamma_i \tilde{d}_{it}(\omega)V_{it}.$$ 

Before we conclude this section, we would like to point out that since the probability distributions of random variables associated with $o_{it}(\omega)$ can depend on the equipment’s processing speed, this leads to a stochastic programming model with decision-dependent uncertainty. When the relationship between the failure probability distribution and the equipment processing speed can be explicitly characterized properly, e.g., based on analysis using extensive experiment results, we can resort to a (nonlinear) stochastic programming
formulation that incorporates the decision-dependent uncertainty:

\[
\min h^\top I_0 + \sum_{s=1}^{S} p_s(V) f(X_s, I_s, \omega_s)
\]

s.t. (3.3), (3.7),

\[
A'X_s = 0, \quad s = 1, 2, \ldots, S,
\]
\[
X_{ts} = b(V_t, I_0, I_{ts}, \omega_s), \quad \forall t \in \mathcal{T}, s = 1, 2, \ldots, S,
\]
\[
X_{ts} \leq d(V_t, \omega_s), \quad \forall t \in \mathcal{T}, s = 1, 2, \ldots, S,
\]
\[
\underline{\iota}(\omega_s) \leq I_{ts} \leq \bar{\iota}(\omega_s), \quad \forall t \in \mathcal{T}, s = 1, 2, \ldots, S,
\]
\[
\sum_{s=1}^{S} p_s(V) \mathbb{1}[R(X_s, \omega_s) \geq r] \geq 1 - \hat{\epsilon},
\]

where \(p_s(V)\) characterizes the dependency of the probability of occurrence of each scenario \(s = 1, 2, \ldots, S\) on decision variables \(\{V_t\}_{t \in \mathcal{T}}\). However, currently we do not have enough experimental data to adequately model this dependency. We leave this as a future research direction to pursue.

3.4 Case Study

3.4.1 Data Collection

We develop a case study using historical data and the design of the biomass processing system from the PDU. The operational costs are collected from the Biomass Logistics Model (BLM), which is an integrated software framework that simulates the entire supply chain and calculates associated costs, energy consumption and GHG emissions (Cafferty et al., 2013). The summary table of the operational costs is presented in Appendix A.

Our data set summarizes the data related to the processing of 14 bales of switchgrass during a period of 4 days at PDU that includes the amount of biomass processed, equipment throughput, electrical current and power consumption per time period (of 0.2 seconds). The moisture level of biomass during each time period is recorded via sensors. Changes of moisture level and dry matter losses during grinding and pelleting operations, and system stoppages are also recorded.

We use another data set from the PDU that includes the density of biomass and particle size distribution measured at three different points in this process: (i) before the process begins while biomass is still
in bale format, (ii) after biomass is processed in the first grinder, and (iii) after biomass is processed in the second grinder. The PDU uses a screen of size 76.2mm (3 inch) in the primary grinder and screens of size 6.35mm (1\(\frac{1}{4}\) inch) in the separation process and the secondary grinder in order to separate particles based on their length. This data is considered in modeling of the separation process.

Work by (Hansen et al., 2019) provides additional data about bulk density of switchgrass when different harvesting equipment are used. They report densities which vary from 171.26 to 234.81 kg/m\(^3\).

Finally, we use data generated by the DEM model to create a regression model that represents the relationship between moisture content, particle size distribution and bulk density. Tables A11 to A14 in the Appendix summarize the input data used. We have consulted the experts and operators of the PDU during model development, verification and validation.

### 3.4.2 Data Analysis

We assume that the density of biomass bales follows a triangular distribution. This distribution is typically used when the number of samples is small and conducting additional sampling is expensive (Kotz and van Dorp, 2004; Thomopoulos, 2017), which is the case here.

The moisture level of bales in our case study varies from 5% to 25%, and bales are grouped into low (5% to 10%), medium (10% to 17.5%) and high (17.5% to 25%) moisture levels. We fitted a uniform distribution to describe the distribution of moisture in each level.

The following regression models are developed using data from the DEM model. The DEM model assumes that biomass particle size follows a uniform distribution. Regression (3.19) presents biomass density after processed at grinder 1. Regression (3.20) presents biomass density after processed at grinder 2. These regressions estimate density as a function of moisture level and particle size. The error term \(\xi_{it}\) is normally distributed, \(\xi_{it} \sim N(0, 3.106)\) and \(\xi_{it}^2 \sim N(0, 10.783)\).

\[
\tilde{d}_{it} = 56.183 + 65.312\tilde{m}_{it} - 8.473\rho_{50} + 0.015\frac{\rho_{90}}{\rho_{10}} + \xi_{it}^1, \quad (3.19)
\]

\[
\tilde{d}_{it} = 186.348 + 206.1697\tilde{m}_{it} - 110.302\rho_{50} + 0.709\frac{\rho_{90}}{\rho_{10}} + \xi_{it}^2. \quad (3.20)
\]

Table A17 in Appendix E summarizes the statistical analysis of regression functions (3.19) and (3.20). The \(R^2\) values for these regression functions are over 94%. Based on the P-values found, the impact of \(\frac{\rho_{90}}{\rho_{50}}\) on biomass density is statistically insignificant. Thus, we do not include \(\frac{\rho_{90}}{\rho_{50}}\) in regression functions (3.19) and
The time-to-failure and duration of an equipment failure depend on the moisture level of biomass. We use historical data from the PDU to estimate these parameters. The duration of a failure is modeled using the uniform distribution, and time-to-failure is modeled using the Weibull distribution. The corresponding parameters for these distributions are summarized in Table 3.2.

<table>
<thead>
<tr>
<th>Moisture Level</th>
<th>Shape</th>
<th>Scale (sec)</th>
<th>Min. (sec)</th>
<th>Max. (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1.16</td>
<td>9.94</td>
<td>0.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Medium</td>
<td>0.83</td>
<td>15.09</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>High</td>
<td>0.59</td>
<td>22.91</td>
<td>0.0</td>
<td>12.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moisture Level</th>
<th>Shape</th>
<th>Scale (sec)</th>
<th>Min. (sec)</th>
<th>Max. (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>5.50</td>
<td>95</td>
<td>20.0</td>
<td>35.0</td>
</tr>
<tr>
<td>Medium</td>
<td>5.00</td>
<td>90</td>
<td>20.0</td>
<td>45.0</td>
</tr>
<tr>
<td>High</td>
<td>4.50</td>
<td>85</td>
<td>20.0</td>
<td>60.0</td>
</tr>
</tbody>
</table>

### 3.5 Numerical Experiments and Sensitivity Analysis

The goal of our numerical experiments is two fold. First, in Section 3.5.1 we evaluate the performance of the model and algorithms proposed. Next, in Sections 3.5.2 to 3.5.6 we evaluate the performance of the system. The performance of the system is measured via (i) reactor utilization, (ii) operational cost, and (iii) inventory level. Section 3.5.2 summarizes the results of our base-case scenario and Sections 3.5.3 to 3.5.6 summarize the sensitivity analysis with respect to bale sequencing, equipment failure rate, storage capacity, and biomass characteristics, respectively.

The algorithms proposed are programmed in Julia 0.7.0 using modeling language JuMP (Dunning et al., 2017). These models ran on Clemson University’s high performance resource Palmetto Cluster and used 16 nodes and 96 GB RAM. The linear and mixed integer programs are solved using the commercial solver GUROBI.

#### 3.5.1 Evaluating Model and Algorithm Performance

In our numerical experiments, we present solutions of SAA model ($\hat{P}$) under different settings. We solve ($\hat{P}$) using a bisection search-based heuristic approach (see Algorithm 3 in Appendix D). To find an appropriate sample size $S$ for the SAA model ($\hat{P}$), we conducted the following stability test. We first solved model ($\hat{P}$) by varying the number of scenarios. Each scenario represents a sample path realization of random biomass characteristics ($\omega$) over the planning horizon. For a given number of scenarios, we ran 10 replications
and computed the relative difference between the objective values associated with these replications. Our experiments showed that with 100 scenarios, the relative difference between the maximum and minimum objective values among the 10 replications is only 0.3%, indicating that the sample size of 100 is appropriate. We thereby use a sample of size $S = 100$ scenarios in our numerical experiments.

To evaluate the performance of the solutions found via the SAA model ($\hat{P}$), we consider an out-of-sample evaluation procedure for the first-stage solution $V^*$ and $I_0^*$. We independently generate 10,000 scenarios in the out-of-sample test and check if solution $(V^*$ and $I_0^*)$ satisfies the chance constraint (3.8) with these 10,000 scenarios. Specifically, we compute the flow values $X_s$ according to equation (3.10) using the first-stage solution $V^*$ and $I_0^*$ for each out-of-sample scenario $s$, and then we check if the reactor utilization target is achieved for at least $(1 - \epsilon) \times 100$ percent of the 10000 scenarios.

Notice that, constraints (3.2)-(3.7) of (P) define system capacities and thresholds that need to be satisfied under all possible realizations of the random variables. Although in the SAA model, the corresponding constraints (3.9) to (3.12) are satisfied for the scenarios generated, enforcing these constraints only for the scenarios used by the SAA model may not be sufficient to ensure that the obtained first-stage solution $V^*$ and $I_0^*$ is feasible under any possible scenario. To ensure that the first-stage solution to the SAA model is “absolutely” feasible under any scenario, one may need to resort to a robust optimization approach by modeling the entire support of random variables $\omega$ as an uncertainty set. This will result in additional (deterministic) constraints in the SAA model ($\hat{P}$), however, we do not expect that this will affect the analysis below. We shared more details for the robust optimization approach in Appendix B.3.

We next demonstrate the value of the proposed chance-constrained stochastic programming model. First, we compare the solution of ($\hat{P}$) with that of the mean-value (MV) problem based on their out-of-sample performance. The MV problem considers a single deterministic scenario where all the random parameters in the problem are replaced by their mean value for the entire duration of the planning horizon. Our experiment result shows that the MV solution ($V$ and $I_0$) led to infeasible solutions in all 10,000 scenarios of the out-of-sample evaluation, and the amount of violation in the metering bin capacity constraint was as much as 666% in some of the scenarios. This indicates that it is necessary to use a stochastic programming model like the proposed model (P) and its SAA model ($\hat{P}$) to adequately address the stochasticity in the problem.

We next justify the use of the proposed chance-constrained stochastic program to model system’s reliability, which is defined as the probability of achieving the desired target feeding rate (see constraint (3.8)). Based on our discussions with the PDU operator, we set the reliability level to be 90%. We conducted a sensitivity analysis to evaluate the impact of the established reliability level on system’s performance. The
results of this analysis are summarized in Table 3.4 and Figure 3.2. These experiments assume that the biomass processed in the system has low moisture level. Similar observations are made when the system processes biomass with higher moisture level. Table 3.3 and Figure 3.2 illustrates the trade-off between costs, target reactor feeding rate, average inventory, maximum inventory and $\epsilon$. As reliability level decreases, the operational cost, the average inventory and the maximum inventory decrease and a higher reactor target rate is achieved.

Finally, Table 3.4 summarizes the numerical results from solving the MIP formulation ($\bar{P}$) and from solving formulation ($\bar{\bar{P}}$) using Algorithm 1 for the same problem instances listed in Table 3.3. These results indicate that bisection search algorithm provides solutions of the same quality as the MIP formulation. Additionally, the running time of the bisection search algorithm is 0.6 to 2.9 times faster.

### Table 3.3: System’s Performance vs. Reliability

<table>
<thead>
<tr>
<th>System Reliability (%)</th>
<th>Reactor Target (dt/hr)</th>
<th>Average Inventory (tons)</th>
<th>Maximum Inventory (tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>3.68</td>
<td>0.56</td>
<td>0.88</td>
</tr>
<tr>
<td>99</td>
<td>3.76</td>
<td>0.56</td>
<td>0.87</td>
</tr>
<tr>
<td>90</td>
<td>3.79</td>
<td>0.54</td>
<td>0.84</td>
</tr>
<tr>
<td>80</td>
<td>3.81</td>
<td>0.52</td>
<td>0.81</td>
</tr>
</tbody>
</table>

### Figure 3.2: Total Cost vs. Reliability

### Table 3.4: Summary of Results from MIP Formulation and Algorithm 1

<table>
<thead>
<tr>
<th>Algorithm 1 Run Time (sec.)</th>
<th>MIP Model Run Time (sec.)</th>
<th>Error Gap (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.9</td>
<td>30.8</td>
<td>0.0</td>
</tr>
<tr>
<td>7.4</td>
<td>23.4</td>
<td>0.0</td>
</tr>
<tr>
<td>9.4</td>
<td>22.3</td>
<td>0.0</td>
</tr>
<tr>
<td>10.2</td>
<td>16.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

* Error Gap = (MIP Obj - Alg. 1 Obj)/(MIP Obj)*100

### 3.5.2 Base Case Analysis

In this section, we summarize our experimental results for a base-case problem, which considers that every biomass bales, each of the same moisture level, is processed in a system with no equipment failures. We consider that storage capacity is 49.1 m$^3$, the particle size follows a uniform distribution, and the risk level is set to 90%. Additional problem parameters are presented in Appendix C.

Tables 3.5 and 3.6 summarize the performance of the system under two moisture levels and two
reactor capacities. Throughout this section, we consider 2.7 dt/hr to be low capacity, and 4.8 dt/hr to be high. We observe that reactor utilization is higher and the operational cost is lower when processing low moisture biomass since the in-feed rate is higher. The inefficiency of the system when operating high moisture biomass lead to low reactor utilization. For example, when reactor’s capacity is high, processing high moisture biomass leads to only 44% utilization of the reactor.

Table 3.5: Base-Case Problem: Reactor Utilization & Costs

<table>
<thead>
<tr>
<th>Bale Moisture Level</th>
<th>Reactor Capacity (dt/hr)</th>
<th>Reactor Flow (dt/hr)</th>
<th>Reactor Utilization (%)</th>
<th>Energy Cost ($/dt)</th>
<th>Fixed Costs ($/dt)</th>
<th>Total Cost ($/dt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>2.25</td>
<td>83</td>
<td>3.85</td>
<td>46.60</td>
<td>50.45</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>3.85</td>
<td>80</td>
<td>2.25</td>
<td>27.26</td>
<td>29.54</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>2.10</td>
<td>77</td>
<td>6.96</td>
<td>50.33</td>
<td>57.29</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>2.11</td>
<td>44</td>
<td>6.93</td>
<td>50.13</td>
<td>57.06</td>
</tr>
</tbody>
</table>

The results of Table 3.6 indicate that, when processing biomass with high moisture level, the initial inventory level is higher and infeed rate is lower than when processing biomass with low moisture level. The additional inventory is needed to ensure a continuous flow of biomass to the reactor when moisture level is high. The corresponding average and maximum inventory are also higher.

Table 3.6: Base-Case Problem: Inventory Level & Equipment Setting

<table>
<thead>
<tr>
<th>Bale Moisture Level</th>
<th>Reactor Capacity (dt/hr)</th>
<th>Reactor Flow (dt/hr)</th>
<th>Initial Inventory (tons)</th>
<th>Average Inventory (tons)</th>
<th>Maximum Inventory (tons)</th>
<th>Average System Infeed Rate (inch/min)</th>
<th>Metering Bin Conveyor Speed (inch/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>2.25</td>
<td>0.23</td>
<td>0.37</td>
<td>0.53</td>
<td>6.7</td>
<td>4.0</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>3.85</td>
<td>0.25</td>
<td>0.56</td>
<td>0.88</td>
<td>11.5</td>
<td>6.9</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>2.10</td>
<td>0.21</td>
<td>0.71</td>
<td>2.12</td>
<td>5.5</td>
<td>4.4</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>2.11</td>
<td>0.28</td>
<td>0.54</td>
<td>2.18</td>
<td>5.5</td>
<td>4.5</td>
</tr>
</tbody>
</table>

The results of Tables 3.5 and 3.6 provide the best and worst performance of the system since we consider that biomass moisture level is either low or high. Section 3.5.3 evaluates system’s performance when a mix of bales of different moisture level are processed.

3.5.3 Sensitivity Analysis: Bale Sequencing

In this section, we analyze the impact of bale sequencing on the performance of the system. Sequencing is a commonly used strategy to improve the reliability of a system (Coit and Zio, 2018). Since the moisture level impacts the processing time of a bale (Zhou et al., 2008), (Crawford et al., 2016), bales are sequenced according to moisture level. We consider that 60% of the bales processed are of low moisture, 10% are of medium moisture, and 30% are of high moisture. We sequence bales based on moisture level. The following sequences are considered, long, short and random. The long sequence considers that every
bale of a particular moisture level is processed before the processing of bales of another moisture level. In our experiments, the long sequence begins by processing high moisture bales, then medium moisture bales and ends with low moisture bales. Such a sequence leads to the worst performance among all long sequences because it requires higher initial inventory to maintain a continuous feeding of the reactor while processing high moisture bales. We consider this specific sequence to highlight the difference in the performance of long versus short sequences.

The short sequence follows a pattern of 60% low, 10% medium, and 30% high moisture bales. This pattern repeats itself several times during the planning horizon. Finally, a random sequence processes bales of different moisture level using a random pattern of high, medium and low moisture. The results presented in Table 3.7 are the averages from 10 problem generated using these sequencing approaches.

Table 3.7: Bale Sequencing: Results of the Sensitivity Analysis

<table>
<thead>
<tr>
<th>Sequencing Approach</th>
<th>Reactor Capacity (dt/hr)</th>
<th>Reactor Flow (dt/hr)</th>
<th>Reactor Utilization (%)</th>
<th>Energy Cost ($/dt)</th>
<th>Fixed Costs ($/dt)</th>
<th>Total Cost ($/dt)</th>
<th>Average Inventory (tons)</th>
<th>Max Inventory (tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td>Low</td>
<td>2.31</td>
<td>85</td>
<td>4.60</td>
<td>45.60</td>
<td>50.20</td>
<td>0.58</td>
<td>1.71</td>
</tr>
<tr>
<td>Long</td>
<td>High</td>
<td>3.39</td>
<td>71</td>
<td>3.13</td>
<td>31.14</td>
<td>34.26</td>
<td>0.72</td>
<td>2.07</td>
</tr>
<tr>
<td>Short</td>
<td>Low</td>
<td>2.30</td>
<td>77</td>
<td>4.60</td>
<td>45.62</td>
<td>50.23</td>
<td>0.41</td>
<td>0.61</td>
</tr>
<tr>
<td>Short</td>
<td>High</td>
<td>3.68</td>
<td>77</td>
<td>2.88</td>
<td>28.61</td>
<td>31.49</td>
<td>1.05</td>
<td>1.76</td>
</tr>
<tr>
<td>Random</td>
<td>Low</td>
<td>2.31</td>
<td>85</td>
<td>4.60</td>
<td>45.61</td>
<td>50.21</td>
<td>0.41</td>
<td>0.63</td>
</tr>
<tr>
<td>Random</td>
<td>High</td>
<td>3.62</td>
<td>75</td>
<td>2.93</td>
<td>29.17</td>
<td>32.11</td>
<td>0.84</td>
<td>1.92</td>
</tr>
</tbody>
</table>

Table 3.7 summarizes the results of the sensitivity analysis with respect to bale sequencing. Short sequences outperform long and random sequences. Short sequences perform best when the capacity of the reactor is high. In this case, the flow to the reactor is highest at 3.68 dt/hr, utilization is highest at 77%, and total cost is lowest at $31.49/dt.

Figure 3.3 presents the inventory level of each sequence. In both Figures/Chapter3 3.3a and Fig-
ure 3.3b, the inventory level for short sequences has a cyclic pattern since inventory accumulates at a high rate while processing low moisture bales, inventory accumulates at a low rate while processing medium moisture bales, and inventory is used while processing high moisture bales. Long sequences begin with the highest level of initial inventory in order to maintain a continuous flow of biomass to the reactor since these sequences begin with high moisture bales. The performance of the random sequence is similar to the short sequence when the processing capacity of the reactor is low. The short sequence outperforms the random sequence when the processing capacity of the reactor is high. The short sequence leads to about 2% lower total costs. Nevertheless, our model does not consider the cost of sequencing the bales. Thus, if the costs of creating a short sequence is high, a biomass processing plant should ensure that bales are processed at some random patterns. The plant should avoid the use of long sequences.

3.5.4 Sensitivity Analysis: Equipment Failure

In this section, we analyze the impact of short-duration and long duration equipment failure on the performance of the system.

**Short-duration Failures:** Recall that, short-duration failures are due to the overfeeding of the system, overflowing of conveyors, or overheating of the grinders. These failures last for no more than 15 seconds, after which the equipment begins working automatically. Table 3.8 summarizes the impact of short-duration failures on reactor utilization, total cost and inventory level.

A comparison of the results of Table 3.8 with Tables 3.5 and 3.6, indicates that short duration failures do not impact reactor’s utilization and total cost when every bale has low moisture. However, the average inventory level increases by 13.6% and maximum inventory level increases by 15.9%. These increase of inventory enables the system to maintain a continuous flow of biomass to the reactor. The performance of the system deteriorates most when every bale processed has high moisture level. The decrease in reactor utilization is 4% and the increase in costs is 4.2%.

**Long-duration Failures:** Recall that, long-duration failures are due to clogging which typically happens when biomass with high moisture level and large particle size is processed. Table 3.9 summarizes the impact of long-duration failures on reactor utilization, total cost and inventory level. The frequency and duration of clogging depends on the moisture level of biomass being processed (Figure 3.2). Thus, every problem in this table faces a different frequency and duration of long failures.

A comparison of the results of Tables 3.7 and 3.9 indicates that long duration failures greatly impact the performance of the system. The utilization of the reactor is reduced by 5 to 58%. The total cost is
Table 3.8: System Performance under Short-duration Equipment Stoppages

<table>
<thead>
<tr>
<th>Biomass Feeding Pattern</th>
<th>Reactor Capacity</th>
<th>Reactor Total Utilization (%)</th>
<th>Reactor Total Cost ($/dt)</th>
<th>Initial Average Maximum Inventory</th>
<th>Average Inventory (tons)</th>
<th>Maximum Inventory (tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>83</td>
<td>50.49</td>
<td>0.23</td>
<td>0.42</td>
<td>0.62</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>80</td>
<td>29.54</td>
<td>0.71</td>
<td>0.61</td>
<td>0.94</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>74</td>
<td>59.69</td>
<td>2.15</td>
<td>1.10</td>
<td>2.15</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>42</td>
<td>59.41</td>
<td>2.15</td>
<td>0.64</td>
<td>2.15</td>
</tr>
<tr>
<td>Long</td>
<td>Low</td>
<td>84</td>
<td>50.59</td>
<td>2.20</td>
<td>0.68</td>
<td>2.20</td>
</tr>
<tr>
<td>Long</td>
<td>High</td>
<td>70</td>
<td>54.72</td>
<td>2.20</td>
<td>0.72</td>
<td>2.20</td>
</tr>
<tr>
<td>Short</td>
<td>Low</td>
<td>84</td>
<td>50.74</td>
<td>0.35</td>
<td>0.48</td>
<td>0.76</td>
</tr>
<tr>
<td>Short</td>
<td>High</td>
<td>75</td>
<td>32.38</td>
<td>2.37</td>
<td>0.98</td>
<td>2.37</td>
</tr>
<tr>
<td>Random</td>
<td>Low</td>
<td>85</td>
<td>50.22</td>
<td>0.40</td>
<td>0.50</td>
<td>0.79</td>
</tr>
<tr>
<td>Random</td>
<td>High</td>
<td>74</td>
<td>32.73</td>
<td>2.10</td>
<td>0.78</td>
<td>2.10</td>
</tr>
</tbody>
</table>

increased 5 to 140%. The maximum inventory level is reached in every problem in Table 3.9.

These observations raise the question whether an increase of storage capacity would allow for additional accumulation of the inventory which could be processed by the reactor during the time an equipment is down. Section 3.5.5 investigates the impact of increasing storage capacity on system’s performance. We also observe that the performance of the system is worst when every bale has high moisture level. The performance of the system is better when short sequences are processed as compared to long or random sequences.

Notice that the value of the maximum inventory in Table 3.9 differs from one problem to the next for two reasons. First, biomass density is a random variable, thus, the volume that X tons of biomass take is also a random variable. However, the volume of the metering bin is fixed. Thus, the maximum inventory (in tons) stored in the metering bin differs based on biomass density. Second, the results in Table 3.9 present the averages of maximum inventory over 10 replications. Thus, some of the differences observed in these values are due to reporting averages.

Table 3.9: System Performance under Long-duration Equipment Failures

<table>
<thead>
<tr>
<th>Biomass Feeding Pattern</th>
<th>Reactor Capacity</th>
<th>Reactor Total Utilization (%)</th>
<th>Reactor Total Cost ($/dt)</th>
<th>Initial Average Maximum Inventory</th>
<th>Average Inventory (tons)</th>
<th>Maximum Inventory (tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>79</td>
<td>53.00</td>
<td>0.29</td>
<td>1.19</td>
<td>2.03</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>52</td>
<td>45.68</td>
<td>1.26</td>
<td>1.24</td>
<td>2.12</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>44</td>
<td>99.48</td>
<td>2.15</td>
<td>1.41</td>
<td>2.19</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>25</td>
<td>99.37</td>
<td>2.22</td>
<td>1.33</td>
<td>2.22</td>
</tr>
<tr>
<td>Long</td>
<td>Low</td>
<td>61</td>
<td>69.84</td>
<td>2.15</td>
<td>1.40</td>
<td>2.20</td>
</tr>
<tr>
<td>Long</td>
<td>High</td>
<td>38</td>
<td>63.68</td>
<td>2.15</td>
<td>1.43</td>
<td>2.24</td>
</tr>
<tr>
<td>Short</td>
<td>Low</td>
<td>63</td>
<td>67.43</td>
<td>1.44</td>
<td>1.25</td>
<td>2.03</td>
</tr>
<tr>
<td>Short</td>
<td>High</td>
<td>40</td>
<td>59.77</td>
<td>1.63</td>
<td>1.24</td>
<td>2.07</td>
</tr>
<tr>
<td>Random</td>
<td>Low</td>
<td>56</td>
<td>75.95</td>
<td>1.72</td>
<td>1.32</td>
<td>2.10</td>
</tr>
<tr>
<td>Random</td>
<td>High</td>
<td>31</td>
<td>76.81</td>
<td>1.65</td>
<td>1.06</td>
<td>2.01</td>
</tr>
</tbody>
</table>
3.5.5 Sensitivity Analysis: Storage Capacity

In this section, we analyze the impact of an increase of storage capacity on the performance of the system. We consider a 25% and 50% increase of capacity, and resolve (i) the base-case problem, (ii) problems with short-duration equipment failures, and (iii) problems with long-duration equipment failures. Next, we summarize our findings.

**Base-case Problem:** Table 3.10 summarizes the impact of increased storage capacity on costs and reactor utilization for the cases when reactor’s capacity is low and high. The last two columns of this table present the change in these performance measures as compared to the results of the base-case problem (Tables 3.5 and 3.7).

Table 3.10: Base-case Problem: Increased Storage Capacity

<table>
<thead>
<tr>
<th>Biomass Feeding Inventory Pattern</th>
<th>Reactor Total Utilization Cost (Original Cap.)</th>
<th>Reactor Total Utilization Cost (Increased Cap.)</th>
<th>Reactor Total Utilization Cost (Increased Capacity)</th>
<th>Reactor Total Utilization Cost (Original Capacity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Reactor Capacity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>79</td>
<td>56.18</td>
<td>2.1</td>
<td>-1.9</td>
</tr>
<tr>
<td>High</td>
<td>81</td>
<td>54.69</td>
<td>5.0</td>
<td>-4.5</td>
</tr>
<tr>
<td>High Reactor Capacity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>46</td>
<td>54.60</td>
<td>4.7</td>
<td>-4.3</td>
</tr>
<tr>
<td>Long</td>
<td>72</td>
<td>33.75</td>
<td>1.7</td>
<td>-1.5</td>
</tr>
<tr>
<td>Long</td>
<td>73</td>
<td>33.30</td>
<td>3.1</td>
<td>-2.8</td>
</tr>
<tr>
<td>Short</td>
<td>77</td>
<td>31.15</td>
<td>1.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>Short</td>
<td>77</td>
<td>31.15</td>
<td>1.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>Random</td>
<td>76</td>
<td>31.68</td>
<td>1.4</td>
<td>-1.3</td>
</tr>
<tr>
<td>Random</td>
<td>77</td>
<td>31.28</td>
<td>3.1</td>
<td>-2.8</td>
</tr>
</tbody>
</table>

*Δ := (Original Cap. – Increased Cap.)/Original Cap.*

The results of Table 3.10 indicate that a 25% increase of storage capacity led to 1.1 to 2.5% increase of reactor’s utilization and 1.1 to 2.4% decrease of costs. A 50% increase led to 1.1 to 5% increase of reactor’s utilization and 1.1 to 4.5% decrease of costs. The greatest improvements are observed when every bale processed has high moisture level, and when long sequences are processed. In these problems, the maximum inventory reached the storage capacity (see Tables 3.5 and 3.7). Thus, by increasing storage, additional inventory accumulated, which led to increased utilization of the reactor. The problems with short sequences show the least improvements since the maximum inventory level was lower than capacity (see Tables 3.5 and 3.7). Thus, increasing storage capacity has minimal impact on reactor utilization.

**Short-duration Failures:** Table 3.11 summarizes the impact of increased storage capacity on costs and reactor utilization for the cases when reactor’s capacity is low and high. The last two columns of this table present the change in these performance measures as compared to the results of the problem under short
duration equipment failures (Table 3.8).

The results of Table 3.11 indicate that a 25% increase of storage capacity led to 1.2 to 2.7% increase of reactor’s utilization and 1.2 to 2.6% decrease of costs. A 50% increase led to 2.5 to 4.4% increase of reactor’s utilization and 2.3 to 4.0% decrease of costs. The greatest improvements are observed when every bale processed has high moisture level, and when long sequences are processed. The increase of storage capacity has a greater impact on reducing costs and increasing reactor’s utilization as compared to the base-case problem (Table 3.10). This is because the increase of storage capacity allows for additional accumulation of inventory which is used during equipment failures.

Table 3.11: Short-Duration Failures: Increased Storage Capacity

<table>
<thead>
<tr>
<th>Biomass Feeding Pattern</th>
<th>Inventory Capacity Increase (%)</th>
<th>Reactor Total Utilization (%)</th>
<th>Total Cost ($/dt)</th>
<th>Reactor Utilization Δ* (%)</th>
<th>Total Cost Δ* (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low Reactor Capacity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>25</td>
<td>75</td>
<td>38.68</td>
<td>1.8</td>
<td>-1.7</td>
</tr>
<tr>
<td>High</td>
<td>50</td>
<td>78</td>
<td>57.30</td>
<td>4.4</td>
<td>-4.0</td>
</tr>
<tr>
<td><strong>High Reactor Capacity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>25</td>
<td>43</td>
<td>57.89</td>
<td>2.7</td>
<td>-2.6</td>
</tr>
<tr>
<td>High</td>
<td>50</td>
<td>44</td>
<td>57.17</td>
<td>4.1</td>
<td>-3.8</td>
</tr>
<tr>
<td>Long</td>
<td>25</td>
<td>71</td>
<td>34.11</td>
<td>1.9</td>
<td>-1.7</td>
</tr>
<tr>
<td>Long</td>
<td>50</td>
<td>72</td>
<td>33.56</td>
<td>3.7</td>
<td>-3.3</td>
</tr>
<tr>
<td>Short</td>
<td>25</td>
<td>75</td>
<td>32.00</td>
<td>1.2</td>
<td>-1.2</td>
</tr>
<tr>
<td>Short</td>
<td>50</td>
<td>76</td>
<td>31.64</td>
<td>2.5</td>
<td>-2.3</td>
</tr>
<tr>
<td>Random</td>
<td>25</td>
<td>75</td>
<td>32.11</td>
<td>2.0</td>
<td>-1.9</td>
</tr>
<tr>
<td>Random</td>
<td>50</td>
<td>76</td>
<td>31.72</td>
<td>3.4</td>
<td>-3.1</td>
</tr>
</tbody>
</table>

*Δ := (Original Cap. – Increased Cap.)/Original Cap.

Long-duration Failures: Table 3.12 summarizes the impact of increased storage capacity on costs and reactor utilization for the cases when reactor’s capacity is low and high. The last two columns of this table present the change in these performance measures as compared to the results of the problem under long-duration equipment failures (Table 3.9).

The results of Table 3.12 indicate that a 25% increase of storage capacity led to 3.9 to 33.8% increase of reactor’s utilization and 3.7 to 24.9% decrease of costs. A 50% increase led to 3.9 to 51.7% increase of reactor’s utilization and 3.7 to 33.8% decrease of costs. An increase of storage capacity has the greatest impact on reducing costs and increasing reactor utilization when the system experiences long-duration failures. The results of Table 3.9 indicate that the maximum inventory reached storage capacity in all problems solved. Thus, increasing the storage capacity leads to additional accumulation of the inventory which is used to maintain a continuous flow of biomass to the reactor during failures.
Table 3.12: Long-Duration Failures: Increased Storage Capacity

<table>
<thead>
<tr>
<th>Biomass Feeding Pattern</th>
<th>Inventory Capacity Increase (%)</th>
<th>Reactor Total Utilization (%)</th>
<th>Total Cost ($/dt)</th>
<th>Reactor Utilization $\Delta^*$ (%)</th>
<th>Total Cost $\Delta^*$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low Reactor Capacity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>25</td>
<td>82</td>
<td>41.03</td>
<td>3.9</td>
<td>-3.7</td>
</tr>
<tr>
<td>Low</td>
<td>50</td>
<td>82</td>
<td>51.02</td>
<td>3.9</td>
<td>-3.7</td>
</tr>
<tr>
<td>High</td>
<td>25</td>
<td>51</td>
<td>87.02</td>
<td>14.4</td>
<td>-12.5</td>
</tr>
<tr>
<td>High</td>
<td>50</td>
<td>56</td>
<td>79.72</td>
<td>25.0</td>
<td>-19.9</td>
</tr>
<tr>
<td>Long</td>
<td>25</td>
<td>70</td>
<td>61.26</td>
<td>14.1</td>
<td>-12.3</td>
</tr>
<tr>
<td>Long</td>
<td>50</td>
<td>76</td>
<td>56.64</td>
<td>23.6</td>
<td>-18.9</td>
</tr>
<tr>
<td>Short</td>
<td>25</td>
<td>70</td>
<td>60.50</td>
<td>11.4</td>
<td>-10.3</td>
</tr>
<tr>
<td>Short</td>
<td>50</td>
<td>77</td>
<td>55.46</td>
<td>21.6</td>
<td>-17.8</td>
</tr>
<tr>
<td>Random</td>
<td>25</td>
<td>65</td>
<td>65.56</td>
<td>15.8</td>
<td>-13.7</td>
</tr>
<tr>
<td>Random</td>
<td>50</td>
<td>72</td>
<td>59.46</td>
<td>27.7</td>
<td>-21.7</td>
</tr>
<tr>
<td><strong>High Reactor Capacity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>25</td>
<td>61</td>
<td>39.05</td>
<td>17.1</td>
<td>-14.5</td>
</tr>
<tr>
<td>Low</td>
<td>50</td>
<td>68</td>
<td>35.10</td>
<td>30.3</td>
<td>-23.2</td>
</tr>
<tr>
<td>High</td>
<td>25</td>
<td>29</td>
<td>86.32</td>
<td>15.2</td>
<td>-13.1</td>
</tr>
<tr>
<td>High</td>
<td>50</td>
<td>32</td>
<td>78.50</td>
<td>26.8</td>
<td>-21.0</td>
</tr>
<tr>
<td>Long</td>
<td>25</td>
<td>45</td>
<td>54.24</td>
<td>17.6</td>
<td>-14.8</td>
</tr>
<tr>
<td>Long</td>
<td>50</td>
<td>50</td>
<td>48.04</td>
<td>32.8</td>
<td>-24.6</td>
</tr>
<tr>
<td>Short</td>
<td>25</td>
<td>46</td>
<td>52.37</td>
<td>14.2</td>
<td>-12.4</td>
</tr>
<tr>
<td>Short</td>
<td>50</td>
<td>51</td>
<td>47.51</td>
<td>26.0</td>
<td>-20.5</td>
</tr>
<tr>
<td>Random</td>
<td>25</td>
<td>42</td>
<td>57.64</td>
<td>33.8</td>
<td>-24.9</td>
</tr>
<tr>
<td>Random</td>
<td>50</td>
<td>48</td>
<td>50.83</td>
<td>51.7</td>
<td>-33.8</td>
</tr>
</tbody>
</table>

$\Delta := (\text{Original Cap.} - \text{Increased Cap.})/\text{Original Cap.}$

### 3.5.6 Sensitivity Analysis: Biomass Particle Size

In this section, we analyze the impact of particle size on the performance of the system. Regression equations (3.19) and (3.20) show the relationship between particle size distribution (represented by $\rho^{50}$) and biomass density. Biomass density impacts the weight of biomass that can be stored in the metering bin. Particle size also impacts the separation process after the first grinder. Our experiments focus on evaluating the impact that (i) particle size ($\rho^{50}$), and (ii) particle uniformity ($\rho^{10}/\rho^{90}$) have on the performance of the system.

Table A12 in Appendix C summarizes the distribution of particle size for the base-case problem. In the following analysis we use as a reference this problem for the case when low moisture biomass is processed and reactor’s capacity is high. Similar observations are made for problems that use other biomass feeding patterns.

**Primary Grinder:** Table 3.13 summarizes the change of total cost and reactor’s utilization due to increases of particle size. The distribution of particle size can be controlled by changing the rotational speed of mills in the primary grinder. Figure 3.4 summarizes the distribution of the mean particle size in the metering bin,
and the corresponding standard deviation.

One would expect that by increasing particle size, the density of biomass in the metering bin would decrease. Different from this intuition, an initial increase of particle size leads to an increase of biomass density in the metering bin. This is mainly because, a larger proportion of biomass will need to be reprocessed in the secondary grinder to reduce particle size. Reprocessing leads to particles of smaller size (as compared to processing only on the primary grinder) in the metering bin, which leads to increased biomass density and decreased volume. As a result, additional biomass can be stored in the metering bin, which enables the system to maintain a continuous flow of biomass to the reactor. This leads to an increase of reactor’s utilization and a decrease of costs.

However, further increases of particle size decreases biomass density, what leads to an decrease of the weight of biomass flow in the equipment that feed the secondary grinder. Specifically, the weight of biomass that is moved via drag chain (DC 5 & DC6) and screw (SC6) conveyors decreases. For example, when mean particle size increases by 3mm, SC6 can transport 1.0 to 6.2 ton/hr. This limits the flow of biomass to the reactor.

![Figure 3.4: Average and Standard Deviation of Biomass in Metering Bin.](image)

**Table 3.13: Particle Size: Primary Grinder**

<table>
<thead>
<tr>
<th>$\Delta \rho_{50}$ (mm)</th>
<th>Bypass Ratio (%)</th>
<th>$\Delta^*$ Reactor Utilization (%)</th>
<th>$\Delta^*$ Total Cost (%)</th>
<th>D5 Capacity Min (ton/hr)</th>
<th>D5 Capacity Max (ton/hr)</th>
<th>D6 Capacity Min (ton/hr)</th>
<th>D6 Capacity Max (ton/hr)</th>
<th>SC6 Capacity Min (ton/hr)</th>
<th>SC6 Capacity Max (ton/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>87</td>
<td>0.0</td>
<td>0.0</td>
<td>11.5</td>
<td>21.2</td>
<td>11.5</td>
<td>21.2</td>
<td>6.2</td>
<td>11.3</td>
</tr>
<tr>
<td>+1</td>
<td>58</td>
<td>+6.0</td>
<td>-5.6</td>
<td>7.7</td>
<td>17.4</td>
<td>7.7</td>
<td>17.4</td>
<td>4.1</td>
<td>9.3</td>
</tr>
<tr>
<td>+2</td>
<td>53</td>
<td>+4.7</td>
<td>-4.5</td>
<td>5.8</td>
<td>15.4</td>
<td>5.8</td>
<td>15.4</td>
<td>3.1</td>
<td>8.2</td>
</tr>
<tr>
<td>+3</td>
<td>51</td>
<td>-3.1</td>
<td>+3.3</td>
<td>1.9</td>
<td>11.5</td>
<td>1.9</td>
<td>11.5</td>
<td>1.0</td>
<td>6.2</td>
</tr>
</tbody>
</table>

* $\Delta := (Original \ \rho_{50} - Changed \ \rho_{50}) / Original \ \rho_{50}$

Table 3.14 summarizes the results of changing the particle size uniformity on reactor utilization and
Table 3.14: Particle Uniformity: Primary Grinder

<table>
<thead>
<tr>
<th>$\Delta \rho_{90}^{\text{Ratio}}$</th>
<th>Bypass Ratio (%)</th>
<th>$\Delta^* \rho_{10}^{\text{Reactor}}$ Utilization (%)</th>
<th>$\Delta^* \rho_{90}^{\text{Total}}$ Cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>91</td>
<td>-2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>-2</td>
<td>95</td>
<td>-3.4</td>
<td>3.5</td>
</tr>
<tr>
<td>-3</td>
<td>100</td>
<td>-1.8</td>
<td>1.8</td>
</tr>
</tbody>
</table>

$*\Delta := (\text{Original } \rho_{90}^{\text{Ratio}} - \text{Changed } \rho_{90}^{\text{Ratio}})/\text{Original } \rho_{90}^{\text{Ratio}}$

Table 3.15: Particle Size: Secondary Grinder

<table>
<thead>
<tr>
<th>$\Delta \rho_{50}^{\text{util}}$</th>
<th>$\Delta \rho_{50}^{\text{Reactor}}$ Utilization (%)</th>
<th>$\Delta \rho_{50}^{\text{Total}}$ Cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ 0.1</td>
<td>+ 0.8</td>
<td>-0.8</td>
</tr>
<tr>
<td>+ 0.2</td>
<td>+ 1.5</td>
<td>-1.4</td>
</tr>
<tr>
<td>+ 0.3</td>
<td>+ 2.3</td>
<td>-2.3</td>
</tr>
</tbody>
</table>

$*\Delta := (\text{Original } \rho_{50}^{\text{util}} - \text{Changed } \rho_{50}^{\text{util}})/\text{Original } \rho_{50}^{\text{util}}$

costs. Notice that, by decreasing $\frac{\rho_{90}}{\rho_{10}}$, we increase particle size uniformity, which means, the distribution of particle size gets close to the mean value. This leads to an increased bypass ratio, which is calculated via equation (28). Figure 3.5 indicates a decrease of biomass density in the metering bin. These changes lead to decreases of reactor utilization and increases cost.

**Secondary Grinder:** Tables 3.15 summarizes the impact of changes in particle size after the secondary grinder to the total cost and reactor utilization. Changes in particle size and particle uniformity after the secondary grinder do not impact the bypass ratio and the flow of equipment upstream. Thus we did not consider changes in particle uniformity after the secondary grinder. Figure 3.6 presents the decrease of normalized standard deviation of biomass density in the metering bin. Biomass density in the metering bin follows a similar trend. The results show that increase of particle size leads to increased utilization of the reactor and decreased total costs. Since the particle size of biomass about to be processed in the secondary grinder is smaller than that of biomass processed in the primary grinder, we consider smaller changes of particle size. Thus, the resulting changes to reactor utilization and costs are smaller.
Figure 3.5: Average Biomass Density

Figure 3.6: Normalized Standard Deviation of Biomass Density in the Metering Bin
Chapter 4

Adaptive Biomass Processing

Operations under Uncertainty

4.1 Background

The future of manufacturing lies in creating a connected digital system for physical assets (e.g., vehicles, equipment, machines, robots), people, and information systems. Integration of these different elements will result in smart manufacturing facilities and contribute to transitioning into Industry 4.0 with higher productivity, better quality control, and more flexible operations without the interference of a human operator (Frank et al., 2019). One of the important elements in Industry 4.0 is sensing technology (Dalenogare et al., 2018). The use of sensory data allows the integration of sequentially revealed information into the decision-making process and increases the facilities’ adaptability to dynamic system environments such as the uncertainty of characteristics of materials being processed. In this chapter, we focus on biomass processing systems in biorefineries. Biorefineries use monitoring and sensing technologies to measure biomass characteristics before they are processed. Biorefineries typically use sensors to measure moisture level and weight. For example, Snetterton and Sleaford Renewable Energy Plants are two of the several commercial-size biomass processing plants in the UK (2015). Both plants use automated bale handling systems that utilize sensors to collect data about moisture level of incoming bales. The data is used to control their operations.

Bales vary based on the physical/chemical characteristics of biomass (such as ash, moisture, carbohydrate contents, etc.). These variations are due to spatial variations in weather and soil, harvesting equipment
used, etc. For example, consider the scenario when a number of bales of different types of feedstocks with different moisture levels and ash content are processed on the same equipment (e.g., a grinder). The resulting distribution of particle size and particle uniformity of processed biomass vary from one bale to the next. These variations lead to uneven biomass flow in a biomass processing system that could affect equipment utilization rate and lead to inconsistent conversion rates. The system may revise equipment processing speed to mitigate these variations. The sensory data enables these revisions to be adaptive to the changes in biomass' characteristics. Some examples of equipment processing speed are the rotational speed for the conveyors and discharge rate for the storage equipment. The discharge rate for the storage equipment controls the inventory level inside it, which can help “smooth out” the biomass flow inside the biomass feeding system. This helps to achieve the prescribed system utilization target. The objective of this work is to investigate the value of using sensory data and a stochastic optimization model to provide adaptive system control with uncertain biomass characteristics.

We propose a multi-stage stochastic programming (SP) model that minimizes the expected operational cost by identifying an optimal operational policy for equipment speed setting and infeed rate in a biomass processing system. Multi-stage SP is a well-known decision-making framework where the uncertainty (e.g., biomass characteristics) is revealed over time, and decisions are made sequentially based on the information revealed in each stage. Multi-stage SP models extend the two-stage SP models that provide optimal static strategic and tactical decisions by finding optimal adaptive decision policy in each time stage based on the uncertainty realized so far (Birge, 1985a). The policy obtained from the multi-stage SP model will take sensory data and the current biomass inventory level as inputs and provide dynamic control on inventory levels by setting equipment processing speed and infeed rate in an adaptive fashion.

We expect that the results of this study will facilitate the creation of a automated process control inside biorefineries that will increase productivity, result in better quality control and seamless and flexible operations. In addition, the results of this study will help biorefineries to develop strategies that increase reactor utilization and minimize costs; and to identify an optimal operational condition in the face of stochastic biomass characteristics. These outcomes will facilitate the commercial-scale generation of biofuels at a competitive cost. In the long run, these outcomes will strengthen the sustainable bioeconomy of the US, enhance the security of energy supplies, reduce dependencies on fossil fuels, and reduce greenhouse gas (GHG) emissions. A strong and sustainable bioeconomy has additional socio-economic benefits, such as generating new green jobs, growth of rural economy and social stability (Domac et al., 2005; Sims, 2003; You et al., 2012).
The remainder of the chapter is organized as follows. We conduct a brief literature review in Section 4.2. In Section 4.3, we describe the problem setting and the modeling framework. In Section 4.4, we present our solution approach and implementation details. We conduct a case study and summarize numerical results and sensitivity analysis in Section 4.5.

4.2 Literature Review

Our study is closely related to two main streams of literature, namely multi-stage stochastic programming and optimization of biorefinery operations.

4.2.1 Multi-stage Stochastic Programming

Multi-stage SP is a well-known decision-making framework where the uncertain information is revealed over time, and the decisions are made sequentially based on the data available at each stage. Real-world applications of multi-stage SP models are rich in many areas such as energy (Bhattacharya et al., 2018; de Matos et al., 2017; Bruno et al., 2016; Siddig and Song, 2019), finance (Carino et al., 1994; Steinbach, 1999; Gulpinar et al., 2002; Dupačová, 2009), facility capacity planning (Ahmed and Sahinidis, 2003; Chen et al., 2002; Gupta and Grossmann, 2014; Singh et al., 2009), and transportation (Alonso et al., 2000; Herer et al., 2006; Möller et al., 2008). While the main advantage of multi-stage models is to develop adaptive decision policies, it is computationally challenging to solve these models. The challenge is due to the uncertainty in the data and the nested structure of the multi-stage decision-making problem. A typical approach to handle the uncertainty is to approximate the underlying stochastic process using scenario trees (Birge and Louveaux, 1997). However, as the number of decision stages in the planning horizon increases, the size of the scenario tree grows exponentially due to the nested structure (Shapiro and Nemirovski, 2005). Many decomposition algorithms have been proposed in the literature to address this burden. (Birge, 1985b) extends the known two-stage L-shaped or Benders decomposition to multi-stage linear SP problems. The main idea of this nested Benders decomposition algorithm is to develop an outer approximation of the expected cost of the future stages (i.e., cost-to-go functions) by generating Benders cuts. Note that, as the number of stages increases, the nested Benders decomposition may become inefficient to solve the problem. (Pereira and Pinto, 1991) proposed the Stochastic Dual Dynamic Programming (SDDP) algorithm for multi-stage problems where the underlying random vectors are stage-wise independent and the expected cost-to-go functions are convex. Stage-wise independence means the probability distribution of the random variable that
is observed in stage $t$, $\xi_t$, is independent of the realization of $\xi_{t-1}$. The stage-wise independence property allows one to define only a single expected cost-to-go function for every stage. SDDP is an iterative algorithm that has shown to have finite convergence with probability one under mild conditions (Shapiro, 2011), which maintains a cutting plane approximation for the expected cost-to-go functions. In our work, we apply the SDDP algorithm to solve the proposed multi-stage SP model. We share more details about the algorithm in Section 4.4.

The literature shows the value of multi-stage SP (VMS) in many application areas. The VMS measures the relative advantage of multi-stage solutions over their two-stage counterparts. (Huang and Ahmed, 2009) provides analytical bounds for the VMS in capacity planning under uncertainty. Similarly, (Xie and Huang, 2018) shows the VMS for strategic expansion of biofuel supply chain under uncertainty. (Mahmuto˘gulları et al., 2019) presents the VMS in risk-averse unit commitment under uncertainty, a daily problem that arises energy market participation. This chapter contributes to the current multi-stage SP literature by demonstrates the VMS for the optimization of manufacturing and processing facilities.

4.2.2 Optimization of Biorefinery Operations

Biorefinery operations include functions such as biomass preprocessing, biomass handling, and biomass storage. Works that focus on evaluating the design parameters of the equipment used cover most of the literature related to biorefinery operations (Crawford et al., 2016; Dai et al., 2012). Other studies analyze the energy consumption of equipment as a function of its design parameters and biomass characteristics (Jacobson et al., 2014b; Kenney et al., 2014; Yancey and Tumuluru, 2015). These studies are limited in scope since they do not capture the interactions between equipment and the impact of equipment on the system’s performance. In our work, we take a system’s approach and use a network flow model to characterize the flow of biomass over time during the planning.

A number of studies use optimization models for optimal process control. These models minimize the energy consumption of individual equipment. For example, in (Numbi and Xia, 2016), the authors propose a deterministic, non-linear optimization model to determine a crusher’s optimal operating parameters. The model considers the time-to-use electricity tariff to achieve additional cost savings. Work by (Zhang and Xia, 2010) presents an optimal control model for a series of conveyor belts. The authors evaluate two different operational structures: one with fixed conveyor speeds and another with conveyor speeds that are modified over time. The comparison of these models demonstrates the benefits of having the flexibility to adjust the operating conditions of a conveyor. To summarize, optimal control models are complex non-linear programs
and thus they are limited to model only some part of the system. In our work, we impose simplifying assumptions to avoid complicated non-linear and non-convex constraints in the proposed multi-stage SP model.

A few studies present models for optimal design of operations in a biorefinery. (Pham and El-Halwagi, 2012), for example, proposes a two-step strategy to optimizing a biorefinery’s design and operations. This model identifies an optimal biorefinery configuration for a given set of biomass feedstocks and available conversion technologies. The first step is to determine which intermediary chemicals can be produced using the available feedstocks. The second step is a network flow optimization problem that identifies the optimal combination of intermediary chemicals and corresponding conversion technologies that minimize the total production cost. Work by (Zondervan et al., 2011) uses a mixed-integer and non-linear network optimization model to identify an optimal design for a biorefinery. The model identifies process sequences to optimize the production of a set of biofuels and bioproducts. Both works, (Pham and El-Halwagi, 2012) and (Zondervan et al., 2011), use deterministic models since they consider fixed biomass characteristics, flow rates, and yields. Different from these works, our research models the uncertainty of biomass moisture level. In a recent work, we proposed a chance-constrained SP model in (Gulcan et al., 2021) that identifies operating conditions of equipment and inventory level to maintain a continuous flow of biomass to the reactor. This work considers the uncertainty of biomass moisture content, particle size uncertainties, and random equipment failure. This chance-constrained SP model results in static policies that cannot incorporate newly learned information via sensory regarding of biomass characteristics. In this chapter, our proposed multi-stage SP model will enable the integration of the sensory data into the decision-making process and increase the facilities’ adaptability to the uncertain characteristics of biomass.

4.3 Problem Description and Mathematical Formulation

We use a network flow model to characterize the flow of biomass within the system over time during the planning horizon $T$. Recall that, in Chapter 3, we demonstrated the network structure of the system and shared details of the equipment and operation.

The following is a list of notations we use.
SETS:

\( M \) The set of moisture levels of biomass \( M := \{ \text{LOW, MED, HIGH} \} \).

\( T \) The set of stages in the planning horizon. \( T := \{1, 2, \ldots, T\} \).

\( E^p \) The set of processing equipment.

\( E^r \) The set of transportation equipment.

\( E^m \) The set of storage equipment.

\( N = E^p \cup E^r \cup E^m \)

\( \delta^+_i \) The set of equipment that feed storage equipment \( i \in E^m \).

\( \delta^-_i \) The set of equipment that are fed by storage equipment \( i \in E^m \).

PARAMETERS:

\( A \) Node arc incidence matrix of \( G \).

\( A' \) Node arc incidence matrix of \( G'(E^r, A') \).

\( I_0 \) Initial inventory level of the storage equipment (in dt).

\( c^h \) Inventory holding cost for storage equipment (in \$/dt).

\( c^p \) Penalty cost of not satisfying the target utilization of the reactor (in $/h/dt).

\( \kappa_t \) Moisture level of the biomass bales in stage \( t \ (\kappa_t \in M) \).

\( \bar{v}(\kappa_t) \) Upper bound of equipment processing speed, with respect to the moisture level.

\( \bar{i}(\bar{m}_t) \) Inventory holding capacity of the storage equipment (in dt).

\( r \) The target reactor feeding rate (in dt/h).

\( v_{1t}(\kappa_t) \) The system feeding rate, with respect to the moisture level in stage \( t \) (in m/hr).

\( i_r \) Index of the last equipment that feeds the reactor

RANDOM PARAMETERS:

\( \bar{m}_t \) Moisture content of the biomass bales in stage \( t \) (in %)

DECISION VARIABLES:

\( V_t := \{ V_{it} \}_{i \in N} \) Equipment processing speed (in centimeters or rotations per minute) in stage \( t \in T \).

\( I_t := \{ I_{it} \}_{i \in E^m} \) Inventory level in stage \( t \in T \) (in dt).

\( X_t := \{ X_{it} \}_{i \in N} \) Biomass flow in stage \( t \) (in dt).

\( p_t \) Shortfall of biomass to achieve the target feeding of the reactor in stage \( t \in T \) (in dt).

In a multi-stage SP setting, the uncertain information is revealed gradually over time, and the decision-maker may adapt their decisions accordingly. The inventory level \( I_t \) in the storage equipment characterizes the state of the biomass processing system and links stages of the planning horizon. In the context of our problem, sensors measure the moisture content of the biomass, \( \bar{m}_t \), that will be processed over the processing time. Biomass moisture is one of the main factors that impact the biomass flow in the system by affecting biomass density (Zhou et al., 2008), (Crawford et al., 2016).

We use biomass density to calculate the flow of biomass in the system. In addition, transportation and storage equipment have volumetric capacities. The changes in biomass density impact the amount of biomass that a piece of equipment can transport or store. Thus, it is important that the model considers biomass density and how uncertain biomass moisture impacts the density.
Work by (Yancey and Tumuluru, 2015) shows that increase of moisture content results in a higher clogging rate. Thus, it is a common practice to lower processing rates of equipment when processing biomass with high moisture level. As a result, the moisture level of biomass affects both the system infeed rates, $V_t$, and biomass flow, $X_t$. Thus, making decisions by considering the information about the biomass moisture content can improve system’s performance.

The proposed multi-stage SP model is created based on the following key definitions:

- **Stages**: We define a stage as the time period during which the system processes one biomass bale. At each stage, the exact numerical values of moisture content, $\tilde{m}_t$, will be known via sensors’ readings.

- **State variables**: The inventory level ($I_t$) in the storage equipment characterizes the state of the biomass processing system, which is carried over from one stage to the next.

- **Actions**: At each stage, we can change the processing speed of different equipment $V_t$ according to the realization of biomass moisture level (provided by sensors’ readings) and the current inventory levels.

The following assumptions are employed in our model:

- The random variables $\tilde{m}_t$’s are stage-wise independent. This is a realistic assumption in that biomass is stored by bales – the distribution of the random moisture content of the next bale is independent of the realized moisture level of the current bale.

- The system will process a single bale at each stage. This is an assumption we make to clearly model the structure of our problem and define a stage. In Section 4.5.4, we relax this assumption and consider different stage definitions.

- The system infeed rate, i.e., the processing rate of the first equipment, $v_{1t}$, is given as a problem parameter. If $v_{1t}$ is introduced as a decision variable, the problem becomes non-linear and non-convex due to the flow and inventory balance constraints in the storage equipment. The resulting non-convex multi-stage SP model will be computationally much more challenging to solve. In addition, the value of $v_{1t}$ does not impact the total amount of biomass flowing into the system in a stage but only affects the duration of a stage.

- We only consider the uncertainty of moisture content of biomass bales, and we assume that given the moisture content, the biomass density value after being processed in grinders 1 and 2 is calculated via a deterministic function.
• We do not consider the blending of biomass of different moisture content in the storage equipment. As a result, the density of biomass in the storage unit is calculated using the realizations of moisture content of the current stage. Considering the biomass blending inside the storage equipment will lead to a non-convex multi-stage SP model that is very challenging to solve.

We are now ready to present the proposed multi-stage SP model. First, we introduce a generic nested formulation (4.1). Let \( Y_t := \{V_t, X_t, p_t\} \) be the vector of local stage variables (i.e., control variables) for each stage \( t \in \mathcal{T} \), and let \( \mathcal{Y}_t(I_{t-1}, \tilde{m}_t) \) be the feasible region for state variables \( I_t \) and local variables \( Y_t \) given the previous inventory level \( I_{t-1} \) and a realization of random vector \( \tilde{m}_t \), a nested formulation of the multi-stage SP model can be presented as:

\[
\begin{align*}
\min_{(I_1,Y_1) \in \mathcal{Y}_1(I_0,\tilde{m}_1)} & \quad c^h I_1 + c^p Y_1 + \mathbb{E} \left[ \min_{(I_2,Y_2) \in \mathcal{Y}_2(I_1,\tilde{m}_2)} c^h I_2 + c^p Y_2 \\
& \quad + \mathbb{E} \left[ \cdots + \mathbb{E} \left[ \min_{(I_T,Y_T) \in \mathcal{Y}_T(I_{T-1},\tilde{m}_T)} c^h I_T + c^p Y_T \right] \right] \right].
\end{align*}
\]

(4.1)

where the initial realization of the moisture content, \( \tilde{m}_1 \), is assumed to be deterministic.

Unfortunately, formulation (4.1) is computationally intractable due to the nested optimization posed by the sequential nature of the decision-making structure. This challenge can be addressed by using a dynamic programming reformulation. Under the aforementioned stage-wise independence assumption, dynamic programming formulation can be solved efficiently.

Now, we formulate the problem to be solved in each stage \( t \in \mathcal{T} \) using the dynamic programming approach. Given a realization of the random vector \( \tilde{m}_t \) and the previous inventory level \( I_{t-1} \), the optimization problem to be solved in stage \( t \) is given by:
\[ Q_t(I_{t-1}, \tilde{m}_t) := \min c^{h^T} I_t + c^p p_t + Q_{t+1}(I_t) \quad (4.2a) \]

s.t.

\[ A'X_t = 0, \quad (4.2b) \]

\[ X_t = g(V_t, \tilde{m}_t), \quad (4.2c) \]

\[ X_t \leq h(V_t, \tilde{m}_t), \quad (4.2d) \]

\[ I_{it} = I_{i,t-1} + \sum_{j \in \delta^+_{i}} X_{j,t} - \sum_{j \in \delta^-_{i}} X_{j,t}, \forall i \in E^m, \quad (4.2e) \]

\[ p_t \geq r - X_{i_r,t}, \quad (4.2f) \]

\[ p_t \geq 0, \quad (4.2g) \]

\[ 0 \leq V_t \leq \bar{v}_{\kappa t}, \quad (4.2h) \]

\[ 0 \leq I_t \leq \bar{i}(\tilde{m}_t), \quad (4.2i) \]

where

\[ Q_{t+1}(I_t) := \mathbb{E}[Q_{t+1}(I_t, \tilde{m}_{t+1})], \forall t \neq T, \]

and \[ Q_{T+1}(I_T) = 0. \]

In formulation (4.2), the objective is to minimize the expected total inventory holding cost and the expected penalty of not achieving the target reactor feeding rate \( r \). The latter objective helps ensure a consistently high utilization of the reactor, as the reactor is the most expensive equipment in the biorefinery. Constraint (4.2b) represents the flow balance for transportation equipment. Constraint (4.2c) calculates biomass flow from storage and processing equipment. This constraint is a function of processing speed, inventory level, and moisture content of the biomass. Constraint (4.2d) represents the upper limit on the amount of biomass flow from equipment \( i \in N \). Functions \( g(\cdot) \) and \( h(\cdot) \) are linear functions with respect to equipment geometry, equipment processing speed, and biomass density, which depend on the biomass moisture content \( \tilde{m}_t \). Below we share examples of \( g(\cdot) \) and \( h(\cdot) \):

\[ X_{it} = \gamma_i d_{it}(\tilde{m}_t)V_{it}. \]
\[ X_{it} \leq \gamma_i d_{it}(\tilde{m}_t)V_{it}. \]

Here, \( \gamma_i \) represents the geometry of equipment \( i \) and \( d_{it}(\tilde{m}_t) \) represents biomass density as a function of moisture content.

The inventory balance constraints (4.2e) link the inventory levels in successive stages. There can be multiple equipment which feed and are fed by the storage equipment (\( \delta^{-}_i \) and \( \delta^{+}_i \), respectively). Constraints (4.2f) and (4.2g) calculate the shortfall of achieving the target reactor feeding rate, which is penalized in the objective. Constraints (4.2h) and (4.2i) set bounds on the processing speed of equipment and amount of inventory stored, respectively.

### 4.3.1 Two-stage Approximations

The computational complexity of multi-stage SPs grows exponentially with the increase of the number of stages (Shapiro and Nemirovski, 2005). Two-stage SP models are often used to approximate multi-stage SP models by making the state variables static over time as opposed to allowing them to be adaptive to dynamically revealed information. To illustrate this approximation, we use the scenario tree in Fig. 4.1 to demonstrate the structure of the multi-stage SP model. Note that due to our assumption of stage-wise independent uncertainty, the scenario tree structure illustrated in the figure is referred to as a “recombining” scenario tree in the literature.

![Scenario Tree Illustration](image)

**Figure 4.1: Scenario tree illustration for multi-stage SP.**

The following setting creates the two-stage approximation:

- In the first stage (prior to any realization of uncertainty), the inventory levels \( I_t \)’s are determined for all \( t \in T \).
- In the second stage, given the inventory levels for all \( t \in T \), the equipment processing speeds \( V_t \) and biomass flow \( X_t \) are determined based on the realization of uncertainty at each \( t \in T \).
Figure 4.2 represents the scenario tree structure for the two-stage approximation. As we set \( \{I_t\}_{t \in T} \) in the first stage, there is no distinction between different sample paths (a sample path is a sequence of realizations of random variables from \( t = 1 \) to \( t = T \)). Note that, \( \{I_t\}_{t \in T} \) corresponds to the state variables that link different stages together. When \( \{I_t\}_{t \in T} \) is given, there is no link between problems at different nodes in the second stage, and therefore, we can decompose the second stage by solving a separate problem at each node in the scenario tree given a first-stage solution. As we enforce \( \{I_t\}_{t \in T} \) to be static decisions, i.e., they are enforced to be identical for all sample paths, this two-stage approximation is a restriction of the multi-stage SP model. The key advantage of the proposed multi-stage SP model is to adapt decisions with respect to newly revealed information at each stage.

Given a set \( S \) of sample paths (obtained via i.i.d samples from \( \{\tilde{m}_t\}_{t \in T} \)), where \( \tilde{m}_t^s \) gives the realization of \( \tilde{m}_t \) on sample path \( s \), the two-stage approximation model is presented below. Note that decision variables with a superscript are second-stage decision variables, while the ones without any superscript are first-stage decision variables. We will compare this approximation model with (4.2) to assess the trade-off.
between the benefit of adaptive decision-making and the computational effort in our numerical experiments.

\[
\min \sum_{t \in T} c^k I_t + \frac{1}{|S|} \sum_{s \in S} \sum_{t \in T} c^p p_t^s 
\]

\[
\text{s.t. } I_t \geq 0, \forall t \in T
\]

\[
A' X_t^s = 0, \forall t \in T, \forall s \in S
\]

\[
X_t^s = g(V_t^s, \tilde{m}_t^s), \forall t \in T, \forall s \in S
\]

\[
X_t^s \leq h(V_t^s, \tilde{m}_t^s), \forall t \in T, \forall s \in S
\]

\[
I_{i,t} = I_{i,t-1} + \sum_{j \in \delta_i^+} X_{j,t}^s - \sum_{l \in \delta_i^-} X_{l,t}^s
\]

\[
\forall i \in E^m, \forall t \in T, \forall s \in S
\]

\[
p_t^s \geq r - X_{i_r,t}^s, \forall t \in T, \forall s \in S
\]

\[
p_t^s \geq 0, \forall t \in T, \forall s \in S
\]

\[
0 \leq V_t^s \leq \tilde{v}_{\kappa_t}, \forall t \in T, \forall s \in S
\]

\[
0 \leq I_t \leq \bar{I}(\tilde{m}_t^s), \forall t \in T, \forall s \in S.
\]

4.3.2 Deterministic Model: Mean-value Problem

Solving stochastic programs can be computationally challenging. Many decision-makers prefer to solve simpler versions of the real-world problems. In this section, we introduce the mean-value (MV) problem, which is deterministic and simple to use. We demonstrate the value of stochastic solutions (VSS) by conducting an out-of-sample evaluation of the solutions to the MV problem and the solutions to stochastic model.
\[
\begin{align*}
\min & \quad \sum_{t \in T} (c^T I_t + c^T p_t) \\
\text{s.t.} & \quad I_t \geq 0, \quad \forall t \in T \\
& \quad A^T X_t = 0, \quad \forall t \in T \\
& \quad X_t = g(V_t, \bar{m}_t), \quad \forall t \in T \\
& \quad X_t \leq h(V_t, \bar{m}_t), \quad \forall t \in T \\
& \quad I_{it} = I_{i,t-1} + \sum_{j \in \delta^+_i} X_{j,t} - \sum_{l \in \delta^-_i} X_{l,t}, \quad \forall i \in E^m, \quad \forall t \in T \\
& \quad p_t \geq r - X_{i,t}, \quad \forall t \in T \\
& \quad p_t \geq 0, \quad \forall t \in T \\
& \quad 0 \leq V_t \leq \bar{v}_t, \quad \forall t \in T \\
& \quad 0 \leq I_t \leq \bar{\iota}(\bar{m}_t), \quad \forall t \in T.
\end{align*}
\]

4.4 Solution Methodology: Stochastic Dual Dynamic Programming

In this section, we discuss the solution approaches for solving the proposed multi-stage SP model and implementation details. Recall that the SDDP algorithm proposed by (Pereira and Pinto, 1991) is a popular approach for multi-stage SP with stage-wise independent uncertainty and convex expected cost-to-go functions. Both of these two assumptions are satisfied in model (4.2) as it only involves linear constraints and continuous decision variables. The SDDP algorithm maintains and iteratively updates a Benders-type cutting-plane approximation for the expected cost-to-go function \( Q_t(\cdot) \) until a termination criterion is met. There are two main steps in each iteration of the SDDP algorithm: a backward pass and a forward pass.

To illustrate the SDDP algorithm, let us use the following simplified notation for the stage \( t \) problem:

\[
Q_t(I_{t-1}, \xi_t) := \min e^T I_t + Q_{t+1}(I_t) \\
\text{s.t.} \quad A_t I_t = b_t - B_t I_{t-1},
\]
where $\xi_t := (c, A_t, b_t, B_t)$ represents the (random) problem data. In general, some or all of the data $(c, A_t, b_t, B_t)$ can be subject to uncertainty. In the context of our problem, $c$ is deterministic and $(A_t, b_t, B_t)$ are subject to uncertainty according to their dependence on $\tilde{m}_t$.

**Forward pass**  Sample $M << N$ paths, where $N$ is the total number of sample paths in the scenario tree. Based on the current approximation of the expected cost-to-go functions, $\hat{Q}_{t+1}(\cdot)$, and the candidate solution for the state variable in the previous stage, $\bar{I}_{t-1}$, solve the stage $t$ problem to obtain a candidate solution, $\bar{I}_t$, for each stage $t \in \mathcal{T}$ on each of the $M$ sample paths.

**Backward pass**  Let $N_t$ be the number of realizations of the random variable $\xi_t$ considered in the scenario tree at stage $t$. Let $\xi_j^t := (c, A_j^t, b_j^t, B_j^t)$ for $j \in \{1, \ldots, N_t\}$ correspond to realization of the $j^{th}$ node in stage $t$. In a backward pass, the stage $t$ problem, shown below, is solved for all $N_t$ realizations of $\xi_t$:

$$
Q_t(\bar{I}_{t-1}, \xi_j^t) := \min c^\top \bar{I}_t + \hat{Q}_{t+1}(\bar{I}_t)
$$

$$
s.t. A_j^t \bar{I}_t = b_j^t - B_j^t \bar{I}_{t-1} - B_j^t \bar{I}_t - B_j^t \bar{I}_{t-1} (\pi_j^t),
$$

where $\pi_j^t$ is the associated optimal dual vector. After solving all $N_t$ subproblems, we have the following Benders cut:

$$
\theta_{t-1} \geq \frac{1}{N_t} \sum_{j=1}^{N_t} Q_t(\bar{I}_{t-1}, \xi_j^t) - \frac{1}{N_t} \sum_{j=1}^{N_t} B_j^t \pi_j^t (\bar{I}_{t-1} - \bar{I}_{t-1})
$$

This cut is added to the collection of cutting planes to improve the approximation $\hat{Q}_t(\cdot)$.

**Termination condition**  As each of the expected cost-to-go function approximation $\{\hat{Q}_t(\cdot)\}_{t=2}^{T}$ is a lower approximation to the true function, at each iteration, the optimal value of the first-stage problem with $\hat{Q}_2(\cdot)$ provides a deterministic lower bound. After each forward pass, we can calculate a statistical upper bound for the problem. For each sample path $j = 1, \ldots, M$, let $v_j$ be the sum of immediate cost associated with the trained optimal policy over all stages. Let $\bar{v} := \frac{1}{M} \sum_{j=1}^{M} v_j$ be the sample average and $\sigma_v^2 := \frac{1}{M-1} \sum_{j=1}^{M} (v_j - \bar{v})^2$ be the sample variance. The statistical upper bound is calculated as follows:

$$
v^{UB} := \bar{v} + \frac{z_\alpha \sigma_v}{\sqrt{M}}
$$

73
Here $z_{\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal distribution, and $v^{UB}$ gives an upper bound for the optimal value of the true problem with a confidence of $(1-\alpha)$. We used $\alpha = 0.025$, which corresponds to the 95% confidence interval. One meaningful termination condition is checking the gap between $\bar{v}$ and $v^{UB}$ and comparing it to a prescribed threshold parameter (Shapiro, 2011). Another termination condition that might be computationally more efficient in practice is checking the progress of the improvement of deterministic lower bound over iterations. The algorithm can be terminated if this lower bound fails to improve by more than a user-specified threshold $\epsilon$ for more than a certain number of consecutive iterations $\sigma$. For example, let $\epsilon = 10^{-4}$ and $\sigma = 10$, if for the last 10 iterations, the lower bound does not improve by more than $10^{-4}$, then the SDDP algorithm terminates.

SDDP Implementation We used the Julia package (Dowson and Kapelevich, 2020) that implements the SDDP algorithm to solve the multi-stage SP model. We considered 500 realizations of $\tilde{\mu}_t$ in each stage $t$ to solve the multi-stage SP model via the SDDP algorithm. The package refers to a solution of the model as the trained policy. Various termination conditions are implemented in this Julia package. We chose to use the bound stalling with a tolerance of $10^{-4}$ and 50 previous iterations to check. Once a policy is obtained by solving the multi-stage SP model, we evaluate this policy using 500 out-of-sample scenarios. In the SDDP.jl package, this procedure is called policy simulation. Specifically, we chose the “SDDP.InSampleMonteCarlo” simulation scheme. SDDP.jl package considers this simulation scheme as in-sample because it considers the same random variable distribution as the training procedure.

The two-stage approximation we considered is a restriction of the multi-stage SP model. Finally, we create out-of-sample validation scenarios to test the performance of decision policies obtained from both models.

4.5 Numerical Experiments and Sensitivity Analysis

The goal of our numerical experiments is twofold: (i) to demonstrate the value of multi-stage SP with respect to their two-stage counterpart and the MV problem; (ii) to evaluate the impact of different modeling and operational choices on the system performance. Specifically, we conduct sensitivity analysis regarding the granularity of decision stages, initial inventory level, length of the bale sequencing pattern, and the ordering in the bale sequence.

We implemented every optimization model considered in the chapter in Julia 1.6.1 using the math-
ematical optimization modeling package JuMP (Dunning et al., 2017). We conducted the experiments on Clemson University’s high-performance computing cluster, the Palmetto Cluster, and used 16 nodes with 32 GB RAM. We solve the multi-stage SP model using the SDDP.jl package (Dowson and Kapelevich, 2020). SDDP.jl is a Julia package for solving multi-stage convex stochastic programming problems using SDDP. We use the commercial solver Gurobi as the optimization solver within the package.

4.5.1 Data Description

We develop a case study using historical data about the characteristics of biomass and the performance of equipment of PDU (Gulcan et al., 2021). We also consulted the experts and operators of the PDU during model development, verification, and validation. Our data set summarizes the sensors reading of moisture content collected during the processing of switchgrass bales at PDU. The moisture content of bales in our case study varies from 3% to 30%, and bales are grouped into low (3% to 12%), medium (12% to 20%), and high (20% to 30%) moisture levels. We assume that the probability distribution of moisture content within each level follows a uniform distribution.

Work by (Hansen et al., 2019) provides additional data about the bulk density of switchgrass when different harvesting equipment are used. The average of the dry biomass densities that they report is 203.04 kg/m$^3$. Note that even though we use the average dry density value, the wet density and the biomass flow vary due to the random moisture content.

Finally, we use data generated by the Discrete Element Method (DEM) model to create regression functions (4.5) and (4.6). These equations represent the relationship between moisture content, particle size distribution, and bulk density. Regressions (4.5) and (4.6) present biomass density after processed at grinder 1 and grinder 2, respectively.

\[
\tilde{d}_{g_1,t} = 56.183 + 65.312\tilde{m}_{g_1,t} - 8.473\tilde{\rho}_{50}^{g_1}
\]  

(4.5)

\[
\tilde{d}_{g_2,t} = 186.348 + 206.1697\tilde{m}_{g_2,t} - 110.302\tilde{\rho}_{50}^{g_2}
\]  

(4.6)

where $d_{it}$ represents the biomass density (dependent variable), $m_{it}$ represents the moisture level of the biomass processed in equipment $i = g_1, g_2$ (indices $g_1$ and $g_2$ correspond to grinder 1 and grinder 2, respectively) and $\tilde{\rho}_{50}^{g_i}$ represents the 50-th percentile of the particle size distribution of biomass processed in equipment $i$. The coefficients $\{\alpha_{0}^{i}, \alpha_{1}^{i}, \alpha_{2}^{i}\}_{i=g_1,g_2}$ in the regression function are determined based on the DEM by (Guo et al.,
that simulates the process in grinders 1 and 2. Tables A12 to A13 in Appendix B.5 summarize the input data used.

4.5.2 Selection of Sample Sizes for SP Models

We first conducted in-sample and out-of-sample stability tests to identify the appropriate number of scenarios to use in the policy training and policy evaluation phases for the SP models under consideration. Let $N_t$ be the number of realizations used in each stage for solving the multi-stage SP model via the SDDP algorithm, and let $S_S$ be the number of sample paths used for solving the two-stage SP model. Additionally, we let $S_V$ be the number of sample paths created to test the performance of any decision policy (provided by either the multi-stage SP, two-stage SP or MV problem) in the out-of-sample evaluation phase.

Multi-stage SP Model - $N_t$ Table 4.1 summarizes the optimal objective value information from the multi-stage SP model with the number of realizations per stage ($N_t$) in the training sample ranging from 250 to 750. Recall that the SDDP algorithm gives a deterministic lower bound (LB) and a statistical upper bound (CI) associated with the obtained decision policy is given by an out-of-sample experiment. We used 500 out-of-sample scenarios in these experiments. The results of Table 4.1 show that there is no significant difference in the out-of-sample performance of the policy for different values of $N_t$ between 250 and 750. Thus, we chose the middle value and set $N_t = 500$ for our experiments.

Two-stage SP Model - $S_S$ Table 4.2 presents the optimal objective value of the two-stage approximation with the number of (training) scenarios $S_S$ ranging from 500 to 1,500. Table 4.2 presents the in-sample performance of the two-stage model. Again, we did not observe any significant difference in the optimal objective value for the values of $S_S$ within the range between 500 and 1,500. Thus, we set $S_S = 1,000$ for our experiments.

<table>
<thead>
<tr>
<th>Number of Realizations $N_t$</th>
<th>250</th>
<th>500*</th>
<th>750</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI* ($$)</td>
<td>[124.26 - 124.55]</td>
<td>[124.41 - 124.69]</td>
<td>[124.38 - 124.66]</td>
</tr>
<tr>
<td>LB ($$)</td>
<td>124.36</td>
<td>124.48</td>
<td>124.45</td>
</tr>
</tbody>
</table>

*95% Confidence Interval
Table 4.2: In-sample stability test - # of scenarios used in the two-stage SP model.

<table>
<thead>
<tr>
<th>Number of Scenarios $S_V$</th>
<th>Objective ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>128.25</td>
</tr>
<tr>
<td>1,000</td>
<td>128.28</td>
</tr>
<tr>
<td>1,500</td>
<td>128.32</td>
</tr>
</tbody>
</table>

**Out-of-sample Evaluation - $S_V$**  
Table 4.3 presents the CI on the optimal cost for different models: multi-stage SP, two-stage SP, and a deterministic model, under different number of ($S_V$) out-of-sample scenarios (ranging from 250 to 750). For these experiments, we use decision policies trained by multi-stage SP and two-stage SP models with $N_t = 500$ and $S_S = 1,000$, respectively. Based on the out-of-sample stability tests, we set $S_V = 500$ for our experiments.

Table 4.3: Out-of-sample stability test - # of sample paths used for performance evaluation.

<table>
<thead>
<tr>
<th>Model</th>
<th>$S_V$</th>
<th>CI* ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-stage</td>
<td>250</td>
<td>[124.28 – 124.68]</td>
</tr>
<tr>
<td>MV Problem</td>
<td></td>
<td>[130.09 – 130.59]</td>
</tr>
<tr>
<td>Two-stage</td>
<td></td>
<td>[128.10 – 128.47]</td>
</tr>
<tr>
<td>Multi-stage</td>
<td>500*</td>
<td>[124.41 – 124.69]</td>
</tr>
<tr>
<td>MV Problem</td>
<td></td>
<td>[130.21 – 130.56]</td>
</tr>
<tr>
<td>Two-stage</td>
<td></td>
<td>[128.22 – 128.48]</td>
</tr>
<tr>
<td>Multi-stage</td>
<td>750</td>
<td>[124.41 – 124.64]</td>
</tr>
<tr>
<td>MV Problem</td>
<td></td>
<td>[130.22 – 130.53]</td>
</tr>
<tr>
<td>Two-stage</td>
<td></td>
<td>[128.23 – 128.44]</td>
</tr>
</tbody>
</table>

*95% Confidence Interval

4.5.3 Performance of the Multi-stage SP Model

In this section, we demonstrate the value of the proposed multi-stage SP model by comparing it with the two-stage approximation and the deterministic MV problem that we presented in Sections 4.3.1 and 4.3.2, respectively.

4.5.3.1 Base-case problem

First, we define a base-case problem, which considers a planning horizon of processing 50 bales, among which 60% are low-moisture, 20% are medium-moisture, and 20% are high-moisture bales. We assume that the bales are sequenced based on a repeating pattern of one high moisture bale, one medium moisture bale, and three low moisture bales over the planning horizon. We call this type of sequencing pattern the “short” sequence. In the base-case problem, we set the target reactor feeding rate to be 2.95 $dt/hr$.

Table 4.4 summarizes the performance of the three models under consideration for the base-case
problem. The results of Table 4.4 demonstrate the value of the proposed multi-stage SP model: it provides
an operational decision policy that costs 3.1% less than the static policy obtained by the two-stage SP model,
and 4.7% less than the static policy obtained by the MV problem, in terms of the total expected cost. Not
surprisingly, this superiority of solution quality comes with a price of excessive computational time. Also, the
CIs of these models do not intersect. Thus difference between solutions provided by each policy is statistically
significant.

As shown in Figure 4.3, the operational policy obtained using the multi-stage SP model performs
better in terms of achieving the target reactor rate by adaptively setting the equipment processing speed
based on the changes in the moisture content measured by the sensors.

Table 4.4: Model performance on the base-case problem.

<table>
<thead>
<tr>
<th>Model</th>
<th>CI* ($)</th>
<th>Gap** (%)</th>
<th>Run Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-stage</td>
<td>[124.41 − 124.69]</td>
<td>-</td>
<td>644</td>
</tr>
<tr>
<td>MV Problem</td>
<td>[130.21 − 130.56]</td>
<td>4.7</td>
<td>7</td>
</tr>
<tr>
<td>Two-stage</td>
<td>[128.22 − 128.48]</td>
<td>3.1</td>
<td>74</td>
</tr>
</tbody>
</table>

*95% Confidence Interval
**Gap := Percentage difference compared to the multi-stage SP using the mean performance (middle point of the CI).

Figure 4.3: Target violation penalties for base-case problem.

Figure 4.4 shows the inventory levels given by the multi-stage SP model for stages 36 to 50. In
Figure 4.4, the solid line represents the average inventory level over 500 sample paths (for the out-of-sample
evaluation) in each stage. We use box-and-whisker plots to represent the fluctuations in the inventory levels
for different sample paths in each stage. As we see from Figure 4.4, the inventory level changes in a cyclic pattern every five stages where the moisture level in the bale sequence completes a whole pattern of one high, one medium, and three low. The inventory is accumulated while processing low-moisture bales because the system can run faster. The system uses the accumulated inventory while processing high- and medium-moisture bales. The decrease in the inventory level is steeper when high moisture bales are processed because the system runs slower and requires higher inventory to meet the target reactor rate. In the last three stages, we observe no inventory due to the end-of-horizon effect.

Table 4.5 compares inventory levels set by the two-stage SP and MV problem with the average inventory level from the multi-stage SP solutions. We see that the two-stage SP model results in lower inventory levels, which reduces the inventory holding costs, but in return, increases the penalty cost caused by the violation of the reactor’s target rate, as shown in Figure 4.3.

Figure 4.4: Percentage difference in the inventory level for the two-stage SP and the MV problem for the base-case problem compared with the multi-stage SP model.

<table>
<thead>
<tr>
<th>Stages</th>
<th>36</th>
<th>37</th>
<th>38</th>
<th>39</th>
<th>40</th>
<th>41</th>
<th>42</th>
<th>43</th>
<th>44</th>
<th>45</th>
<th>46</th>
<th>47</th>
<th>48</th>
<th>49</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-stage</td>
<td>-0.1%</td>
<td>0.0%</td>
<td>-3.6%</td>
<td>-3.1%</td>
<td>-2.7%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>-3.7%</td>
<td>-3.3%</td>
<td>-3.1%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>MV Problem</td>
<td>0.4%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>-0.2%</td>
<td>0.6%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table 4.5: Percentage difference in the inventory level obtained by the two-stage SP model and MV problem in stage 36 to stage 50 for the base-case problem compared to the multi-stage SP model.
4.5.3.2 Length of planning horizon

The goal of the following experiments is to assess the value of the multi-stage SP as the number of stages increases. In our problem setting, the length of the planning horizon is identified by the number of bales that the system will process during the operation time. Table 4.6 summarizes the performance comparison of the three models on problems with different planning horizons. For a fair comparison, Table 4.6 records the total cost per bale. In addition to the base-case problem of processing 50 bales, we considered problems with a planning horizon of 10, 25, and 100 stages/bales.

Table 4.6: Model performances for different length of planning horizons.

<table>
<thead>
<tr>
<th>Model</th>
<th>Cost per Bale CI* ($)</th>
<th>Gap** (%)</th>
<th>Run Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Bales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-stage</td>
<td>3.389 – 3.399</td>
<td>-</td>
<td>128</td>
</tr>
<tr>
<td>MV Problem</td>
<td>3.453 – 3.465</td>
<td>1.9</td>
<td>7</td>
</tr>
<tr>
<td>Two-stage</td>
<td>3.431 – 3.440</td>
<td>1.2</td>
<td>18</td>
</tr>
<tr>
<td>25 Bales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-stage</td>
<td>2.711 – 2.719</td>
<td>-</td>
<td>300</td>
</tr>
<tr>
<td>MV Problem</td>
<td>2.816 – 2.826</td>
<td>3.9</td>
<td>7</td>
</tr>
<tr>
<td>Two-stage</td>
<td>2.779 – 2.786</td>
<td>2.5</td>
<td>38</td>
</tr>
<tr>
<td>100 Bales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-stage</td>
<td>2.375 – 2.380</td>
<td>-</td>
<td>2641</td>
</tr>
<tr>
<td>MV Problem</td>
<td>2.498 – 2.503</td>
<td>5.2</td>
<td>11</td>
</tr>
<tr>
<td>Two-stage</td>
<td>2.456 – 2.460</td>
<td>3.4</td>
<td>299</td>
</tr>
</tbody>
</table>

*95% Confidence Interval
**Gap := Percentage difference compared to the multi-stage SP using the mean performance (middle point of the CI).

The results of Tables 4.4 and 4.6 indicate that as the number of stages increases, the value of using a multi-stage stochastic program increases. Our model proposes that bales are processed based on a sequence that repeats itself. Thus, one might consider using a “truncate-and-repeat” strategy by obtaining an optimal policy from solving a problem with a small number of stages and then repeatedly using the resulting policy as a heuristic approach over the entire planning horizon. Table 4.7 summarizes the outcome of applying this strategy. We can see that for our problem, this “truncate-and-repeat” strategy results in a worse performance that is statistically significant compared to the policy obtained from solving the full-length problem. This is attributed to the “end-of-horizon” effect in multi-stage SP problems (Shapiro, 2011). If the time is limited and the decision needs to be made quickly, rather than using this “truncate-and-repeat” strategy one can use the MV problem or two-stage model. Tables 4.4 and 4.6 show that both models provide higher-quality solutions.
than the solutions obtained by the “truncate-and-repeat” strategy and require much shorter computational
time.

Table 4.7: The performance of the “truncate-and-repeat” strategy.

<table>
<thead>
<tr>
<th></th>
<th>5 Times 10-Bale Policy</th>
<th>2 Times 25-Bale Policy</th>
<th>1 Time 50-Bale Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Cost ($)</strong></td>
<td>[169.5 - 169.9]</td>
<td>[135.6 - 135.9]</td>
<td>[124.4 - 124.7]</td>
</tr>
<tr>
<td><strong>Cost per Bale ($)</strong></td>
<td>[3.389 - 3.399]</td>
<td>[2.711 - 2.719]</td>
<td>[2.488 - 2.494]</td>
</tr>
<tr>
<td><strong>Gap</strong> (%)</td>
<td>36</td>
<td>9</td>
<td>-</td>
</tr>
</tbody>
</table>

*Gap := Percentage difference compared to base-case of 50 bales using the mean performance (middle point of the CI).

4.5.3.3 The mixing ratio of biomass bales based on their moisture levels

The mix of biomass bales based on their moisture levels can greatly impact the performance of
biomass processing system (Gulcan et al., 2021). The goal of the following experiments is to analyze how the
value of multi-stage SP varies under different mixes of biomass bales. In addition to the base-case problem
of processing 60% low - 20% med - 20% high moisture bales, we considered processing all low, all high, 60%
med - 20% low - 20% high and 60% high - 20% med - 20% low moisture bales. For the bale mixes 60% med
- 20% low - 20% high and 60% high - 20% med - 20% low, we considered the bales are sequenced using the
short sequencing strategy with internal patterns of one high moisture bale, three medium moisture bales, and
one low moisture bale, and three high moisture bales, one medium moisture bale, and one low moisture bale,
respectively. For each mix, we experimented with three different target reactor rates; 2.50 $dt/hr$, 2.72 $dt/hr$,
and 2.95 $dt/hr$. We considered that the remainder of the problem parameters is the same as the base-case
problem.
Tables 4.8, 4.9, and 4.10 summarize the results of processing different mixing ratios under different target reactor rates. Tables 4.8, 4.9, and 4.10 show that operational costs increase as the ratio of higher moisture level bales increases in the mix. The reason is that when the system processes high moisture bales, equipment processing rates are lowered to prevent equipment clogging.

The implications of Tables 4.8, 4.9, and 4.10 on the value of multi-stage SP are two-fold. First,
Table 4.10: Model performances for different mixing ratios of bales with target rate 2.50 dt/hr.

<table>
<thead>
<tr>
<th>All Low</th>
<th>60% Low — 20% Med — 20% High</th>
<th>60% Med — 20% Low — 20% High</th>
<th>60% High — 20% Med — 20% Low</th>
<th>All High</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td><em><em>CI</em> ($)</em>*</td>
<td><strong>Gap</strong> (%)</td>
<td><em><em>CI</em> ($)</em>*</td>
<td><strong>Gap</strong> (%)</td>
</tr>
<tr>
<td>Multi-stage</td>
<td>[0.00 – 0.00]</td>
<td>-</td>
<td>[43.22 – 43.31]</td>
<td>-</td>
</tr>
<tr>
<td>MV Problem</td>
<td>[0.00 – 0.00]</td>
<td>0.0</td>
<td>[45.64 – 45.86]</td>
<td>5.7</td>
</tr>
<tr>
<td>Two-stage</td>
<td>[0.00 – 0.00]</td>
<td>0.0</td>
<td>[43.79 – 43.86]</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>60% Low</strong></td>
<td>[68.08 – 68.20]</td>
<td>-</td>
<td>[68.84 – 69.00]</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>60% Med</strong></td>
<td>[70.27 – 70.49]</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>60% High</strong></td>
<td>[284.99 – 285.16]</td>
<td>-</td>
<td>[288.16 – 288.34]</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>All High</strong></td>
<td>[616.77 – 616.95]</td>
<td>-</td>
<td>[616.77 – 616.95]</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*95% Confidence Interval
**Gap := Average change in objective values compared to multi-stage SP model.
when every bale has the same moisture level, there is no value to use a stochastic program to determine the inventory level. This is because the trivial solution of zero inventory is the optimal solution for these cases. When every bale has low moisture level, the target can be achieved without requiring an initial inventory and variation of realizations of moisture content only requires updates to the local control variables (equipment processing rates). Again in the case when moisture level is high, the best strategy is to maintain high feeding rates of the reactor. There are no expectations to accumulate inventory when these bales are being processed. The infeed is the slowest under high moisture bales. Thus, if there is not enough inventory, the target reactor feeding rate cannot be reached. The costs presented in Tables 4.8, 4.9, and 4.10 incur solely due to the penalty for not meeting the targeted feeding rate.

Second, when the system processes a mixture of bales, then, using a multi-stage SP model to optimize the biorefinery operations becomes valuable. The confidence intervals for these different models do not overlap, showing that the differences in different models’ performance are statistically significant. Results of these tables show that different bale mix distributions and different target rates can lead to different values of multi-stage SP.

4.5.4 Sensitivity Analysis: Granularity of Decision Stages

In this section, we analyze the trade-offs among the granularity of decision stages, policy performance, and computational efforts. One of the modeling decisions that impacts the level of adaptability is the granularity of decision stages in the multi-stage SP. Recall that we have assumed thus far that the system processes a single bale at each stage. We now relax this assumption and consider two alternative stage definitions: (i) “Combined” stage definition, where a stage corresponds to the period in which the system will consecutively process biomass bales of the same moisture level; and (ii) “Detailed” stage definition, where a stage is the period in which the system processes one-third of a biomass bale. We assume that these new stage definitions do not conflict with the stage-wise independence assumption, and apply them on the base-case problem presented in Section 4.5.3.

For these experiments, we update the objective function (4.2a) of stage $t$ problem as follows:

$$\min \beta_t (c^T I_t + c^p p_t) + Q_{t+1} I_t,$$

where $\beta_t > 0$ is the number or fraction of bales that the system will process in stage $t$. The presence of $\beta_t$ ensures a unified total cost calculation under different stage definitions.
The “Combined” stage definition is motivated by the argument that we can use a similar operational decision policy for processing biomass bales of the same moisture level. On the other hand, the “Detailed” stage definition aims to interact with the system more frequently to increase adaptability to varying moisture content within the same bale. However, this increases the number of stages and thus increases the solution time for the multi-stage SP model. Table 4.11 summarizes the results of these two alternative stage definitions and compares them with the original stage definition.

Table 4.11: Sensitivity analysis for granularity of decision stages.

<table>
<thead>
<tr>
<th>Stage Definition</th>
<th>CI* ($)</th>
<th>Obj Change** (%)</th>
<th>Run Time (sec)</th>
<th>Run Time Change** (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>[166.09 – 166.38]</td>
<td>33.5</td>
<td>461</td>
<td>-28.4</td>
</tr>
<tr>
<td>Detailed</td>
<td>[123.21 – 123.36]</td>
<td>-1.0</td>
<td>2046</td>
<td>217.7</td>
</tr>
</tbody>
</table>

*95% Confidence Interval
**Change := Percentage change compared to the base-case with multi-stage SP in Table 4.4 using the mean performance (middle point of CI).

The results of Table 4.11 show that the value of increasing the adaptability (with the detailed stage definition) is marginal in terms of solution quality (1% improvement is given by the “Detailed” stage definition). However, this marginal gain is achieved at the expense of a 217.7% increase in computational time. On the other hand, the “Combined” stage definition resulted in a 28.4% reduction in computational time at the cost of a 33.5% deterioration in the solution quality.

4.5.5 Sensitivity Analysis: Initial Inventory Level

In this section, we analyze the impact of the initial inventory level on costs and reactor utilization (i.e., target rate). Inventory holding is a common strategy to ensure high system reliability. We consider three different initial inventory levels for the storage equipment: empty, half-full, and full capacity, according to the common practice implemented at the PDU. For each of these initial inventory levels, we conducted experiments with different target reactor rates.

Figure 4.5 shows how the average operating cost changes as the target reactor rate increases from 1.0 \( \text{dt/hr} \) to 4.0 \( \text{dt/hr} \) for the three initial inventory levels considered in our sensitivity analysis, using the base-case problem described in Section 4.5.3. Note that the total cost calculation only considers the total inventory holding cost accumulated over the planning horizon but does not include the cost for the initial inventory.

In Figure 4.5, we observe that for smaller target rates, starting the operation with empty storage
equipment is more economical. Since the system can achieve the target reactor rate without additional inventory and the system incurs inventory holding cost for not used inventory. However, as the target reactor feeding rate exceeds $2.95 - 3.00 \text{ dt/hr}$, having a half-full initial inventory becomes more economical. The additional initial inventory makes it easier to achieve the target reactor feeding rate. We also see that operating with a full initial inventory is always more costly than the other two strategies for the range of target reactor rates considered.

Another interesting observation from these experiments was that, when operating with a half or full initial inventory, the static solution obtained from the deterministic mean-value problem is infeasible. In order to satisfy the inventory levels set by this static solution, we found out that the system needs to feed the pelleting mill with a rate higher than the pelleting mill’s infeed capacity. This observation points to the need for developing a SP model.

### 4.5.6 Sensitivity Analysis: Length of the Bale Sequencing Pattern

In this section, we analyze the impact of the sequencing choice on the operational cost and reactor utilization. For this analysis, we introduce two different sequencing strategies: the long sequence and the random sequence. The long sequence considers that the system processes every bale of a particular moisture level before processing bales of another moisture level. The operator forms the random sequence by picking bales randomly. In our experiments, we considered 10 random sequences and presented the average results. We solve the base-case problem described in Section 4.5.3 for the three different sequencing strategies (short, long, and random) with a range of target reactor rates.

Figure 4.6 clearly shows the advantage of the short sequence over the entire range of target reactor rates, and in certain cases, it results in up to a 59% reduction in the total cost compared to alternative
options. This is because the repeating pattern in the short sequence enables the system to accumulate inventory while processing low-moisture bales and to consume the inventory while processing high-moisture bales. The performance of the random sequence lies in between the short and long sequencing strategies. One caveat is that our model does not consider the cost of sequencing the bales. Therefore, if the cost of creating a short sequence (due to additional operations such as sorting) is high, a biomass processing plant may simply use the random sequence as a rule of thumb in their operation.

4.5.7 Sensitivity Analysis: Sequence Order

In this section, we analyze the impact of the sequencing order on the total cost and reactor utilization. In the base-case problem, we use the short sequence that repeats the pattern of processing one high-moisture bale, one medium-moisture bale, and three low-moisture bales. Let us call this ordering the high-start sequence. Next, we will compare this the low-start sequence, which we defined to have the repeating pattern of three low-moisture bales, one medium-moisture bale, and one high-moisture bale. We use the base-case problem described in Section 4.5.3 with the high-start and low-start sequences for a range of target rates for the reactor.
We observe from Figure 4.7 that for lower target rates, starting with high-moisture bales results in a better performance. To help facilitate the understanding of this phenomenon, we show in Figure 4.8 the inventory levels for high-start and low-start strategies in each stage. We see that the inventory level presents a cyclic pattern for both strategies. At the beginning and at the end of the planning horizon, the low-start strategy accumulates inventory that will be used when bales with high moisture level are processed. This results in a 56% higher inventory level on average. As the target rate increases, the low-start strategy performs better.

Figure 4.8: Inventory levels for different bale sequence order for \( r = 2.95 \text{ dt/hr} \).

4.5.8 Key Findings

This section summarizes the take-away messages from our numerical experiments and sensitivity analysis. First, using multi-stage SP to model sequential decisions of biorefinery operations improves reactor utilization and costs compared to alternative models (two-stage SP, deterministic mean-value problem) that use only static solutions. Second, the bale-by-bale definition of stages appears to yield the best performance. A more detailed stage definition results in only a marginal gain at the cost of a 217.7% increase in computational time. The combined stage definition results in a 28.4% reduction in computational time but at the expense of a 33.5% deterioration in the solution quality. Third, filling the storage equipment to its half capacity as an initial inventory level improves the system’s capability to meet the reactor’s target feeding rate and reduces costs for high reactor processing rates. Finally, we observe that sequencing bales based on their moisture levels can lead to as much as 59% cost reduction for certain target reactor rates. The short sequences of low-, medium-, and high-moisture bales performs best.
Chapter 5

Conclusions and Future Research

In this dissertation, we study (i) biomass blending problems in supply chains and (ii) biorefinery operations optimization problems. These problems consider a variety of uncertain biomass characteristics.

In Chapter 2, we propose models that identify blends of biomass materials with different physical or chemical properties to satisfy the requirements of the Thermochemical conversion platform at minimum cost. Models consider requirements such as ash and thermal contents. Meeting these process requirements consistently is a challenge because biomass ash and thermal contents are random and vary by supplier. Nevertheless, these requirements should be met most of the time (i.e., 80-90%) in order to optimize the performance of the conversion process. We model these process requirements using chance constraints.

We consider two problem settings, centralized and decentralized supply chains. Most of the supply chain literature assumes a centralized system where a single decision-maker has full control. However, this is not typically the case in many supply chains. The decentralized setting takes into account that biorefinery and suppliers are independent entities that have their own goals and objectives. We propose a Stackelberg game where the biorefinery is the leader of this game and suppliers are the followers. We model this game using a bilevel optimization model with chance constraints.

We use the SAA to approximate the chance constraints. Chapter 2 proposes an exact approach and a heuristic to solve the bilevel optimization model. We also develop a solution algorithm to solve the mixed-integer programming formulation of the centralized model. The case study developed with data from South Carolina enables the testing of the performance of these algorithms. Our numerical analysis indicates that the proposed SAA Algorithm finds high-quality solutions in a reasonable amount of time.

Approximation methods used in this chapter (SAA, linear approximation, etc.) provide lower bound
for the true problem and result in optimistic cost calculations. However, results of the numerical experiments show that these approximations provide high quality solutions. In addition, lower bounds gathered from these approximations allow us to show the quality of proposed heuristic algorithm for solving decentralized supply chain problem that provides an upper bound for the optimal solution.

The suppliers selected and the blends identified when solving the centralized and decentralized models are different. The decentralized model selects blends that contain mainly SN residues and pine. The centralized model uses mainly pine, SP, and SN residues. The amount of hybrid poplar, mixed residues, and C&D waste in both models increases with demand for biomass. Both models suggest the use of 30-45% SN residues and 0% MSW. MSW is not used due to its high ash content. Comparing the costs of meeting demand in the centralized and decentralized supply chains, one can observe that costs of the centralized model are between 2 and 6% lower. The implications of these results are two-fold. First, it could motivate members of the supply chain to collaborate to reduce supply chain costs. Second, it provides an estimate of the approximation error from assuming a centralized supply chain. Decision-makers can use this result to evaluate the trade-offs between improvements of solution accuracy and (computational) efforts of modeling and solving the decentralized model.

In Chapter 3, we move into biorefinery operations. The main objective of this chapter is to develop analytical models which improve the reliability of biomass feeding system in a biorefinery. We propose a stochastic programming model that identifies operating conditions of equipment and inventory level to maintain a continuous biomass flow to the reactor under uncertainty. The uncertain problem parameters are moisture content, particle size distribution, and equipment failures. The proposed chance constraint ensures that a certain reliability level of the reactor is maintained. We propose an SAA of the chance-constraint problem and develop a bisection search-based heuristic to solve this approximation.

This research makes several contributions to the literature. First, the proposed model is developed using real-life data about biomass pre-processing. The model integrates the outcomes of a DEM model into an optimization model. As a result, we expect that practitioners will find the results of this study useful and applicable.

Second, we make a number of important observations. We observe that (i) biomass characteristics impact reactor utilization and costs. The system performs best when processing biomass of low moisture and performs worst when processing biomass of high moisture content. Changes in particle size distribution also impact the system’s performance by increasing reactor utilization by as much as 6% and reducing costs by as much as 5.6%. We observe that (ii) sequencing bales based on moisture level improves the use of resources,
leading to up to a 2.6% increase of reactor’s utilization and 1.96% reduction of costs. We observed that short sequences of low, medium, and high moisture bales perform best. Processing low and medium moisture bales builds up the inventory required to maintain a continuous biomass flow when processing high moisture bales since their processing times are long. We observe that (iii) short equipment failures reduce reactor’s utilization by up to 4.0%, increase costs by up to 4.2%, and increase inventory level by up to 55.8%. Long duration failures reduce reactor’s utilization by up to 58.1%, increase costs by up to 139.2%, increase average inventory by up to 225.8%. The limitations of storage capacity negatively impacted the performance of the system during long-duration failures. Thus, increasing storage capacity and processing short sequences of low, medium, and high moisture biomass are strategies that can be used to mitigate the negative impacts of equipment failure.

In Chapter 4, we consider a different modeling approach to optimize biorefinery operations and incorporated information revealed by moisture sensors inside the biorefinery. The main objective of this work is to minimize operational costs and improve the reliability and flexibility of a biomass processing system in a biorefinery to changing conditions of the processed biomass. We provide a multi-stage stochastic program that revises operational decisions based on new information revealed by moisture sensors. A penalization scheme ensures consistently high utilization of the reactor. We share a mean-value problem and a two-stage approximation for this problem. Comparing these models justifies the modeling choice and presents the value of multi-stage SP to optimize biomass processing operations under uncertainty. We share sensitivity analysis regarding the granularity of decision stages, initial inventory level, length of the bale sequencing pattern, and sequence order. This research makes several contributions to the literature. First, we justify the modeling choice and present the value of multi-stage SP to optimize biomass processing operations under uncertainty. Second, we develop the proposed model using real-life data about biomass preprocessing. The model enables the integration of sensor technologies into automated system control. As a result, we expect that practitioners will find the results of this study helpful and inspiring for the transitioning into Industry 4.0. Third, we make several important observations. We observe that revising the control decisions every time the system starts processing a new bale provides a significant cost advantage compared to less frequent control schemes. We observe that when sequencing of biomass bales is cheap, practitioners should sequence the bales based on moisture content in patterns that repeat frequently. If sequencing is costly, randomly picking biomass bales may be a good strategy.

We have identified other important problems for future research. One line of future work may seek to develop a holistic supply chain model. This model will integrate a macro-level supply chain model
(e.g., models proposed in Chapter 2) that identifies the supplier of a biorefinery based on the moisture level of biomass, selling price, transportation cost, with an optimization model for biorefinery operations (e.g., models proposed in Chapters 3 and 4) that identifies optimal operational conditions for the biorefinery. The model will capture the trade-offs between the biomass selling price and the cost of processing biomass.

In addition, different application areas might benefit from our decentralized supply chain model. This model can be used in variety of supply chains in the presence of raw material quality uncertainty, such as metal supply chains where quality of the metal ores are stochastic in nature and vary from supplier to supplier.

Furthermore, both biorefinery optimization models proposed in Chapters 3 and 4 assume that a single type of biomass feedstock will be processed and use only the moisture level grouping to sequence biomass bales. Several research directions can stem from relaxing these assumptions. On the one hand, additional biomass characteristics such as ash and carbohydrate content can be included in the sequencing decisions. In this case, the complexity of the problem increases, which would motivate the development of efficient solution algorithms. On the other hand, the proposed models can be extended to consider multiple biomass feedstock types and include the sequencing strategies as decision variables. This will lead to endogenous uncertainty since the biomass characteristics during the process depend on the sequencing decision. This more challenging problem will motivate the development of new approximation schemes and computationally efficient solution algorithms.
Appendices
Appendix A  Stochastic Biomass Blending Problem

A.1 Pseudocode for Algorithms

Algorithm 1 SAA Algorithm

Notation: $\lambda'$ and $\lambda''$ are the lower and upper bounds of $\lambda$; $\mu'$ and $\mu''$ are the lower and upper bounds of $\mu$; $\epsilon$ and $\delta$ be small positive constants.

1: while true do
2:   Set $\lambda \leftarrow \lambda' \pm \lambda''$, $\mu \leftarrow \mu' \pm \mu''$.
3:   Solve model $(\bar{P})$ [Solve model $(\bar{Q})$] using GUROBI [using Heuristic Algorithm].
4:   Let $\tilde{W}_s$ and $\tilde{J}_s$ be the incumbent solution of $W_s$ and $J_s$, respectively.
5:   Set $C_1 \leftarrow 0$, $C_2 \leftarrow 0$.
6:   for $s \in [1, \ldots, N]$ do
7:       if $\tilde{W}_s > 0$ then
8:           $C_1 \leftarrow C_1 + 1$
9:       end if
10:      if $\tilde{J}_s > 0$ then
11:         $C_2 \leftarrow C_2 + 1$
12:     end if
13:   end for
14:   if $C_1 \geq \hat{\beta}N + \epsilon$ then
15:       $\lambda' \leftarrow \frac{\lambda' + \lambda''}{2}$
16:   else if $C_1 \leq \hat{\beta}N - \epsilon$ then
17:       $\lambda'' \leftarrow \frac{\lambda' + \lambda''}{2}$
18:   end if
19:   if $C_2 \geq \hat{\gamma}N + \epsilon$ then
20:       $\mu' \leftarrow \frac{\mu' + \mu''}{2}$
21:   else if $C_2 \leq \hat{\gamma}N - \epsilon$ then
22:       $\mu'' \leftarrow \frac{\mu' + \mu''}{2}$
23:   end if
24:   if $|\lambda - \frac{\lambda' + \lambda''}{2}| \leq \delta$ AND $|\mu - \frac{\mu' + \mu''}{2}| \leq \delta$ then
25:     break.
26:   end if
27: end while
28: Return $\lambda$ and $\mu$ and solution of $(\bar{P})$ [solution of $(\bar{Q})$].
Algorithm 2 Heuristic Algorithm to Solve ($\hat{Q}$) for Fixed $\lambda, \mu$

Step 0: Initialize $p = 1$; let $I_b := I, \forall b \in B$; $X_{ibp}^* = \emptyset; C_b^* = \emptyset$
$Z^* = \infty$; counter = 0

Step 1: If $I_b = \emptyset, \forall b \in B$
  Let $I_b := I, \forall b \in B$; Let $p = p + 1$
End

Find $\overline{\pi}_b = \min_{i \in \mathcal{I}_b} \{t_{ib} + \tau_{ib}\}$
Let $i_b^* = \arg\min_{i \in \mathcal{I}_b} \{t_{ib} + \tau_{ib}\}$

Step 2: Set $\hat{c}_b = \overline{\pi}_b$
For $i \in I$
  If $\hat{c}_b \geq (t_{ib} + \tau_{ib})$
    $\hat{X}_{ibp} = \overline{X}_{ibp}$
  Else $X_{ibp} = 0$
End

Step 3: Solve ($\hat{Q}_L$)

Step 4: If ($\hat{Q}_L$) is not feasible:
  Let $I_b = I_b \setminus i_b^*$ GoTo Step 1
Else if $\hat{Z}_L < Z^*$
  $Z^* = \hat{Z}_L; X_{ibp}^* = \hat{X}_{ibp}; C_b^* = \hat{c}_b$
  Let counter = 1; GoTo Step 1
Else if counter != $\nu$
  counter = counter + 1
  Let $I_b = I_b \setminus i_b^*$ GoTo Step 1
Else
  STOP!

Return $X_{ibp}^*$ and $C_b^*$

A.2 Propositions

PROPOSITION 1: The feasible region of ($\hat{P}$) is convex.

Proof: For a given $i$ and $b$, let $f_{ibp}(X_{ib}) = \lambda_{ibp} + c_{ibp}(X_{ib} - \overline{k}_{ibp})$ for $X_{ib} \in [0, \overline{k}_{ibp}]$ and for all $p = 1, \ldots, \mathcal{P}$. Let function $\hat{f}_{ib}(X_{ib}) = \max_{p \in \mathcal{P}} f_{ibp}(X_{ib})$ for $X_{ib} \in [0, \overline{k}_{ibp}]$. That means, $\hat{f}_{ib}(X_{ib}) \geq \lambda_{ibp} + c_{ibp}(X_{ib} - \overline{k}_{ibp})$ for $X_{ib} \in [0, \overline{k}_{ibp}]$. Function $\hat{f}_{ib}(X_{ib})$ is piecewise linear convex since it is the maximum of set of linear functions.

If $\hat{f}_{ib}(X_{ib})$ is a convex function, then, for some constant $s$, the set of $X_{ib}$’s which satisfies: $\hat{f}_{ib}(X_{ib}) \leq s$ is convex (Luenberger et al., 1984). This implies that constraints (2.17) define a convex set. The feasible region of formulation ($\hat{P}$) is the intersection of linear functions (2.12) to (2.14) and the convex set defined by (2.17), thus, it is convex. \[\square\]

12 warnings

PROPOSITION 2: For each feasible solution of ($\hat{P}$) one can find a feasible solution of ($\hat{P}$), and vice versa.

Proof: Let first show that a feasible solution of ($\hat{P}$) is feasible for ($\hat{P}$).

Let $\hat{X}_{ibp}, \hat{Z}_{ibp}$ for all $i \in I, b \in B, p \in \mathcal{P}$ be a feasible solution of ($\hat{P}$). We will show that such a solution satisfies (2.17). This solution satisfies the rest of the constraints of ($\hat{P}$) since they are the same for ($\hat{P}$).
Let $\tilde{X}_{ib} = \tilde{X}_{ibp}$, where, $\tilde{X}_{ibp} = \max_{p \in \mathcal{P}} \tilde{X}_{ibp}$ and $\tilde{p} = \text{argmax}_{p \in \mathcal{P}} \tilde{X}_{ibp}$. Based on constraints (2.7f), for each $i, b$ there is exactly one $\tilde{X}_{ibp} \geq 0$ for all $p \in \mathcal{P}$, thus, $\tilde{X}_{ib} \geq 0$ and $\tilde{X}_{ib} \in [\tilde{k}_{ibp}, \tilde{k}_{ibp}]$. Let $\tilde{F}_{ib} = \lambda_{ibp} + c_{ibp}(\tilde{X}_{ib} - \tilde{k}_{ibp})$. If $\tilde{X}_{ib} = 0$, then $\tilde{F}_{ib} = 0$. One can easily see that $\tilde{X}_{ib}$ and $\tilde{F}_{ib}$ satisfy (2.17).

Let show that a feasible solution of $(\hat{P})$ is feasible for $(\tilde{P})$.

Let $\tilde{F}_{ib}, \tilde{X}_{ib}$ for all $i \in I, b \in \mathcal{B}$ be a feasible solution of $(\hat{P})$. We will show that such a solution satisfies (2.7c) and (2.7f). Recall that constraints (2.7d) are redundant.

If $\tilde{X}_{ib} > 0$ for some $i \in I, b \in \mathcal{B}$, then, let $\tilde{X}_{ib} \in [\tilde{k}_{ibp}, \tilde{k}_{ibp}]$. Set $X_{ibp} = \tilde{X}_{ib}$, $Z_{ibp} = 1$ and $X_{ibp} = Z_{ibp} = 0$, for $p \in \mathcal{P} \setminus \tilde{p}$. If $\tilde{X}_{ib} = 0$ for some $i \in I, b \in \mathcal{B}$, then, let $X_{ibp} = Z_{ibp} = 0$, for $p \in \mathcal{P}$. One can easily see that $\tilde{X}_{ibp}$ and $\tilde{Z}_{ibp}$ satisfy (2.7c) and (2.7f).

**PROPOSITION 3:** An optimal solution of $(\hat{P})$ is a lower bound of $(\tilde{P})$.

**Proof:** Let $X_{ibp}^\ast, F_{ibp}^\ast$ for all $i \in I, b \in \mathcal{B}$ be the optimal solution of $(\hat{P})$ and let $p^\ast$ represent the cost bracket $X_{ibp}^\ast$ corresponds to. Since this is a minimization problem, $F_{ibp}^\ast = \lambda_{ibp} + c_{ibp}(\tilde{X}_{ib} - \tilde{k}_{ibp})$.

One can use this solution to construct a feasible solution $X_{ibp}^\ast, Z_{ibp}^\ast$ of $(\tilde{P})$ as described in Proposition 2. Let $\Delta$ be the difference in objective function value of $(\hat{P})$ and $(\tilde{P})$. Thus, $\Delta = \sum_{i \in I} \sum_{b \in \mathcal{B}} (c_{ibp} \tilde{k}_{ibp} - \lambda_{ibp}) > 0$. □

**PROPOSITION 4:** The linear relaxation of $(\tilde{Q}_{ibp})$ provides an exact solution.

**Proof:** We prove this by contradiction. Let $(\tilde{Q}_{ibp}^F)$ be the linear programming relaxation of $(\tilde{Q}_{ibp})$ obtained by relaxing (2.21c) as follows:

$$Z_p \leq 1, \quad \forall p \in \mathcal{P}. $$

Let assume that an optimal solution to $(\tilde{Q}_{ibp}^F)$ violates constraints (2.21c). Let assume this solution has $X_{p^1}, X_{p^2}, Z_{p^1}, Z_{p^2} > 0$. This solution also satisfies $Z_{p^1} + Z_{p^2} \leq 1$. Let assume that $X_{p^1} \leq X_{p^2}$, thus, $\bar{k}_1 = \bar{k}_3 \leq \bar{k}_2 \leq \bar{k}_4$. The corresponding objective function value is $\tilde{Z}_{ibp}^F = \tau X_{p^1} + \tau X_{p^2}$. Based on constraints (2.21b), $X_{p^1} = \bar{k}_{p^1} Z_{p^1}$ and $X_{p^2} = \bar{k}_{p^2} Z_{p^2}$. Thus, $\tilde{Z}_{ibp}^F = \tau \bar{k}_{p^1} Z_{p^1} + \tau \bar{k}_{p^2} Z_{p^2}$. This means, $\tilde{Z}_{ibp}^F$ is a convex combination of $\tau \bar{k}_{p^1}$ and $\tau \bar{k}_{p^2}$. Consider the following cases:

(a) If $\tau \bar{k}_{p^1} < \tau \bar{k}_{p^2}$: the value of $\tilde{Z}_{ibp}^F$ can increase by letting $Z_{p^1} = 1$ and $Z_{p^2} = 0$.

(b) If $\tau \bar{k}_{p^1} < \tau \bar{k}_{p^2}$: the value of $\tilde{Z}_{ibp}^F$ can increase by letting $Z_{p^1} = 0$ and $Z_{p^2} = 1$.

(c) If $\tau \bar{k}_{p^1} = \tau \bar{k}_{p^2}$, the value of $\tilde{Z}_{ibp}^F$ remains the same by letting $Z_{p^1} = 1$ and $Z_{p^2} = 0$; or $Z_{p^1} = 0$ and $Z_{p^2} = 1$.

In cases (a) and (b), we can improve the objective function value of $(\tilde{Q}_{ibp}^F)$. This contradicts our initial
assumption that the current solution is optimal. In case (c), a solution which does not violate (2.21e) returns the same objective function value. This proves that the linear relaxation ($\bar{Q}_{fb}$) returns a solution which is optimal for ($\bar{Q}_{fb}$).

**Lemma 1:** In an optimal solution to problem ($\bar{Q}_{fb}$), at most one $X_p > 0$ for $p \in P$. Let $p^*$ be cost bracket for which $X_p > 0$, then, $X_{p^*} = \bar{k}_{p^*}$ and $X_p = 0$ for $p \in P \setminus p^*$.

**Proof:** Since the objective of ($\bar{Q}_{fb}$) is to maximize profits, then, if $c > 0$ for some $p \in P$, the corresponding $X_p$ is a candidate optimal solution. Since the objective function is linear, if $c > 0$, then, $X_p = \bar{k}_p$ maximizes profits. Based on (2.21c), in an optimal solution a single cost bracket is selected. Thus, if max$_{p \in P} c\bar{k}_p > 0$, then, in an optimal solution exactly one $X_p > 0$ and the remaining are 0. If max$_{p \in P} c\bar{k}_p < 0$, then, in an optimal solution $X_p = 0, \forall p \in P$. If max$_{p \in P} c_p = 0$, then, let $p^* = \arg\max_{p \in P} c\bar{k}_p$. Solutions with $X_{p^*} = 0, X_{p^*} = \bar{k}$, or $X_{p^*} = \bar{k}$ and $X_p = 0$ for $p \in P \setminus p^*$ lead to the same objective function value of zero. \qed
A.3 Detail Mathematical Formulations

Model formulations: Model (Q).

\[ (Q) : \quad \min : Z^L = \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} (C_b + f_b) X_{ibp} \quad (1a) \]

s.t.

\[ Pr \left( \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} (\bar{a}_{ib} - \alpha) X_{ibp} \leq 0 \right) \geq 1 - \beta, \quad (1b) \]

\[ Pr \left( \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} \bar{e}_b \bar{h}_{ib} X_{ibp} \geq \tau \right) \geq 1 - \gamma, \quad (1c) \]

\[ \sum_{b \in B} \sum_{p \in P} (C_b - \bar{\tau}_{pb} - t_{ib}) X_{ibp} \geq 0, \quad \forall i \in I, \quad (1d) \]

\[ C_b \geq 0, \quad \forall p \in P \quad (1e) \]

\[ \max : Z^F_i = \sum_{b \in B} \sum_{p \in P} (C_b - \bar{\tau}_{pb} - t_{ib}) X_{ibp}, \quad \forall i \in I, \quad (1f) \]

\[ \bar{k}_{ibp} Z_{ibp} \leq X_{ibp} \leq \bar{K}_{ibp} Z_{ibp}, \quad \forall b \in B, p \in P, \quad (1g) \]

\[ \sum_{p \in P} Z_{ibp} = 1, \quad \forall b \in B, \quad (1h) \]

\[ X_{ibp} \geq 0, \quad \forall b \in B, p \in P, \quad (1i) \]

\[ Z_{ibp} \in \{0, 1\}, \quad \forall b \in B, p \in P. \quad (1j) \]
Model formulations: Model \((\bar{Q})\).

\[\begin{align*}
\text{(Q)}: \quad \min & \quad \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} (C_b + f_b)X_{ibp} + \lambda \sum_{s=1}^{N} W_s + \mu \sum_{s=1}^{N} J_s \\
\text{s.t.} & \\
& \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} (a_{ib} - \alpha)X_{ibp} + \nu_s - W_s = 0, \quad \forall s = 1, \ldots, N, \quad (2a) \\
& \tau - \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} e_b h_{ib} X_{ibp} + \nu_s - J_s = 0, \quad \forall s = 1, \ldots, N, \quad (2b) \\
& \sum_{b \in B} \sum_{p \in P} (C_b - \tau_{bp} - t_{ib})X_{ibp} \geq 0, \quad \forall i \in I, \quad (2c) \\
& C_b \geq 0, \quad \forall b \in B, p \in P \quad (2d) \\
& \max: Z_i^f = \sum_{b \in B} \sum_{p \in P} (C_b - \tau_{bp} - t_{ib})X_{ibp}, \quad \forall i \in I, \quad (2e) \\
& \underline{k}_{ibp} Z_{ibp} \leq X_{ibp} \leq \overline{k}_{ibp} Z_{ibp}, \quad \forall b \in B, p \in P, \quad (2f) \\
& \sum_{p \in P} Z_{ibp} = 1, \quad \forall b \in B, \quad (2g) \\
& X_{ibp} \geq 0, \quad \forall b \in B, p \in P, \quad (2h) \\
& Z_{ibp} \in \{0, 1\}, \quad \forall b \in B, p \in P. \quad (2i)
\end{align*}\]

Model formulations: KKT equations of the follower’s problem.

The followers’ problem is a linear program for fixed values of \(c_{bp}\). We replace (2e)-(2h) with the
corresponding KKT conditions.

Constraints: (2f), (2g), (2h)

\begin{align*}
Z_{ibp} &\leq 1, & \forall i \in I, b \in B, p \in P, \\
(C_b - c_{ib} - t_{ib}) + u_{ib} - v_{ibp} + w_{ibp} - l_{ibp} &= 0 & \forall i \in I, b \in B, p \in P, \\
k_{ibp}v_{ibp} - \bar{k}_{ibp}w_{ibp} - m_{ibp} + k_{ibp} &= 0 & \forall i \in I, b \in B, p \in P, \\
(-X_{ibp} + \bar{k}_{ibp}Z_{ibp})v_{ibp} &= 0 & \forall i \in I, b \in B, p \in P, \\
(X_{ibp} - \bar{k}_{ibp}Z_{ibp})w_{ibp} &= 0 & \forall i \in I, b \in B, p \in P, \\
\sum_{p \in P} (Z_{ibp} - 1)\bar{\gamma}_{ib} &= 0 & \forall i \in I, b \in B, \\
X_{ibp}l_{ibp} &= 0 & \forall i \in I, b \in B, p \in P, \\
Z_{ibp}m_{ibp} &= 0 & \forall i \in I, b \in B, p \in P, \\
(Z_{ibp} - 1)k_{ibp} &= 0 & \forall i \in I, b \in B, p \in P, \\
u_{ib}, \gamma_{ib} &\geq 0 & \forall i \in I, b \in B, \\
v_{ibp}, w_{ibp}, l_{ibp}, m_{ibp}, k_{ibp} &\geq 0 & \forall i \in I, b \in B, p \in P.
\end{align*}

Thus, the single level optimization model is the following.

\[
\min : \sum_{i \in I} \sum_{b \in B} \sum_{p \in P} (C_b + f_b)X_{ibp} + \lambda \sum_{s=1}^{N} W_s + \mu \sum_{s=1}^{N} J_s
\]
\[
\text{s.t. } (2a) - (2d), (2f) - (2h), (3a) - (3k).
\]

A.4 Evaluating the Performance of SAA

A.4.1 Finding a feasible solution for \((P)\):

The SAA literature proposes two approaches to generate feasible solutions for \((P)\) (Luedtke and Ahmed, 2008). The first approach identifies the sample size \(N\) prior to solving \((\bar{P})\). For \(\hat{\beta} < \beta, \hat{\gamma} < \gamma\), and for \(N\) large enough, the feasible region of \((\bar{P})\) is a subset of the feasible region of \((P)\). Thus, a feasible solution of \((\bar{P})\) will be feasible for \((P)\) with high probability as \(N \to \infty\). These a priori estimates yield very large sample size \(N\) which impact the size of \((\bar{P})\) and make it impracticable to solve. The second approach uses a smaller sample size \(N\) to find a solution \(\bar{x}\) of \((\bar{P})\), and then, conducts a a posteriori check to see if
Pr(E^1(\bar{x}, \hat{a}) \leq 0) \geq 1 - \beta, and Pr(E^2(\bar{x}, \bar{h}) \leq 0) \geq 1 - \gamma. We use the second approach and provide details below.

Let assume \( \bar{x} \) is a solution returned by SAA Algorithm. To estimate \( p^1(\bar{x}) = Pr\left(E^1(\bar{x}, \hat{a}) \leq 0\right) \) and \( p^2(\bar{x}) = Pr\left(E^2(\bar{x}, \bar{h}) \leq 0\right) \), we sample iid values of the random problem parameters \( \hat{a}, \bar{h} \)
\((a_1, \ldots, a_N, h_1, \ldots, h_{N'})\). This new sample of size \( N' \) is used to calculate estimating probabilities \( \hat{p}^1_N(\bar{x}) \) and \( \hat{p}^2_{N'}(\bar{x}) \) using equations (2.8) and (2.9). By the law of large numbers, probability distribution of \( p^1(\bar{x}) \) can be approximated reasonably close by a normal distribution with mean \( \hat{p}^1_N(\bar{x}) \) and variance \( \hat{p}^1_N(\bar{x}) (1 - \hat{p}^1_N(\bar{x})) / N' \); and probability distribution of \( p^2(\bar{x}) \) can be approximated reasonably close by a normal distribution with mean \( \hat{p}^2_N(\bar{x}) \) and variance \( \hat{p}^2_N(\bar{x}) (1 - \hat{p}^2_N(\bar{x})) / N' \). Using this approximation, one can define one-sided \((1 - \delta)\)-confidence interval for \( p^1(\bar{x}) \) and \( p^2(\bar{x}) \) as follows (Nemirovski and Shapiro, 2006)

\[
p^1(\bar{x}) \leq U^1_{\delta,N'}(\bar{x}),
\]
\[
p^2(\bar{x}) \leq U^2_{\delta,N'}(\bar{x}),
\]
where \( U^1_{\delta,N'}(\bar{x}) = \hat{p}^1_N(\bar{x}) + \Phi^{-1}(1 - \delta) \sqrt{\hat{p}^1_N(\bar{x}) (1 - \hat{p}^1_N(\bar{x})) / N'} \), \( U^2_{\delta,N'}(\bar{x}) = \hat{p}^2_N(\bar{x}) + \Phi^{-1}(1 - \delta) \sqrt{\hat{p}^2_N(\bar{x}) (1 - \hat{p}^2_N(\bar{x})) / N'} \) and \( \Phi^{-1}(\cdot) \) represents the inverse cdf of standard normal distribution. Finally, in order to check the violation of each chance constraint of \((P)\) we compare the values of \( U^1_{\delta,N'}(\bar{x}) \) with \( \beta \) and \( U^2_{\delta,N'}(\bar{x}) \) with \( \gamma \). A solution \( \bar{x} \) returned by SAA Algorithm is feasible to the true problem \((P)\) at \((1 - \delta)\) confidence level if we have \( U^1_{\delta,N'}(\bar{x}, a) \leq \beta \) and \( U^2_{\delta,N'}(\bar{x}, h) \leq \gamma \).

A.4.2 Finding a lower bound for \((P)\):

The SAA can also be used to compute lower bounds for \((P)\) with high confidence. We use the approach proposed by Nemirovski and Shapiro (2006) to calculate lower bounds. We describe this approach below.

Let \( N \) be the total number of observations in a sample, and let \( M \) be the total number of samples generated. The following is the procedure developed to generate a lower bound.

First, we select values for \( N^1, M^1 \). Calculate:

\[
B\left(\lfloor N^1 \beta \rfloor; \beta, N^1\right) = \sum_{i=0}^{\lfloor N^1 \beta \rfloor} \binom{N^1}{i} \beta^i (1 - \beta)^{N^1 - i},
\]
which is the cdf of binomial distribution and represents the probability that \( E^1(x, a_s) > 0 \) in at most \( \lfloor N^1 \beta \rfloor \).
of the $N^1$ observations made ($s = 1, \ldots, N^1$). Let $\pi_{N^1} = B \left( \lceil N^1 \hat{\beta} \rceil; \beta, N^1 \right)$.

Second, we choose $T^1$ to be the largest number which satisfies the following

$$B \left( T^1 - 1; \pi_{N^1}, M^1 \right) \leq \delta.$$ 

Here, $\delta$ is the probability that we observe at most $T^1 - 1$ successes from the total of $M^1$ samples. The probability of success is $\pi_{N^1}$, and a success is a sample in which at least $\lfloor N^1 \hat{\beta} \rfloor$ of the $N^1$ observations made ($s = 1, \ldots, N^1$) satisfy $E^1(x, a^*) > 0$. We follow a similar procedure to identify $M^2$ and $N^2$.

Third, we generate $M (= \max(M^1, M^2))$ independent samples; and each sample contains $N (= \max(N^1, N^2))$ observations of the random vectors $\tilde{a}_{ib}, \tilde{h}_{ib}$. For each sample, we solve problem ($\bar{P}$) to obtain the corresponding optimal objective function values $\vartheta_{N,m}(m = 1, \ldots, M)$. These values represent iid realizations of the random variable $\vartheta^*$. In order to find lower bounds for $\vartheta^*$, we sort these values in a non-decreasing order, i.e. $\vartheta_{N,(1)} \leq \vartheta_{N,(2)} \leq \ldots \leq \vartheta_{N,(M)}$. It is shown that, the quantity $\vartheta_{N,(T)}$ is a lower bound to $\vartheta^*$ with probability at least $(1 - \delta)$.

### A.4.3 Evaluating the Performance of SAA Algorithm to Solve ($P$):

See Appendix A.4 for details of the procedure we use to generate lower and upper bounds for ($P$) via SAA, using a method developed by (Luedtke and Ahmed, 2008).

To evaluate the effectiveness of the SAA in finding feasible solutions for ($P$), we vary the risk level $\beta, \gamma$ and sample size $N$. Tables A1 and A2 summarize the results for $\hat{\beta} = \hat{\gamma} = 0$ and $M = 10$ replications. In Table A1, $\beta = \gamma = 0.3$, and in Table A2, $\beta = \gamma = 0.2$. For each combination of $\beta, \gamma, N$, we calculate the risk of the generated solution and the cost of the feasible solutions found, i.e., those solutions which have risk less than 0.3 and 0.2 correspondingly. For a given solution $x^*$, the risk is $Pr \{ E^1(x^*, \tilde{a}) \neq 0 \}$ and $Pr \{ E^2(x^*, \tilde{h}) \neq 0 \}$. We also report the corresponding average, minimum, maximum and sample standard deviation ($\sigma$) of the risk and the average run time over the 10 replications. We report the number of feasible solutions found. For these feasible solutions, we report the average, minimum, maximum and standard deviation of costs.

Based on the results from Table A1, the algorithm finds 10 feasible solutions for $N \geq 125$. Based on the results from Table A2, the algorithm finds 10 feasible solutions for $N \geq 250$. The corresponding objective function values increase with $N$.

Tables A3 and A4 summarize the results of SAA Algorithm for $\hat{\beta} = \beta$ and $\hat{\gamma} = \gamma$. In this case, in order to find a feasible solution to model ($P$), we increase the sample size $N$. The costs of solutions found are about
Table A1: Solutions Returned by SAA Algorithm for \((P) (\beta = \gamma = 0.3; \hat{\beta} = \hat{\gamma} = 0.0)\)

<table>
<thead>
<tr>
<th>N</th>
<th>Solution Risk</th>
<th># of Feasible Solutions</th>
<th>Objective Value (in $1,000)</th>
<th>Avg Run Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg</td>
<td>Min</td>
<td>Max</td>
<td>(\sigma)</td>
</tr>
<tr>
<td>50</td>
<td>0.443</td>
<td>0.321</td>
<td>0.651</td>
<td>0.101</td>
</tr>
<tr>
<td>75</td>
<td>0.323</td>
<td>0.230</td>
<td>0.468</td>
<td>0.073</td>
</tr>
<tr>
<td>100</td>
<td>0.274</td>
<td>0.164</td>
<td>0.366</td>
<td>0.057</td>
</tr>
<tr>
<td>125</td>
<td>0.233</td>
<td>0.168</td>
<td>0.294</td>
<td>0.044</td>
</tr>
<tr>
<td>150</td>
<td>0.206</td>
<td>0.161</td>
<td>0.283</td>
<td>0.037</td>
</tr>
<tr>
<td>175</td>
<td>0.164</td>
<td>0.121</td>
<td>0.203</td>
<td>0.030</td>
</tr>
<tr>
<td>200</td>
<td>0.170</td>
<td>0.106</td>
<td>0.230</td>
<td>0.038</td>
</tr>
<tr>
<td>250</td>
<td>0.136</td>
<td>0.099</td>
<td>0.172</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Table A2: Solutions Returned by SAA Algorithm for \((P) (\beta = \gamma = 0.2; \hat{\beta} = \hat{\gamma} = 0.0)\)

<table>
<thead>
<tr>
<th>N</th>
<th>Solution Risk</th>
<th># of Feasible Solutions</th>
<th>Objective Value (in $1,000)</th>
<th>Avg Run Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg</td>
<td>Min</td>
<td>Max</td>
<td>(\sigma)</td>
</tr>
<tr>
<td>50</td>
<td>0.443</td>
<td>0.321</td>
<td>0.651</td>
<td>0.101</td>
</tr>
<tr>
<td>75</td>
<td>0.323</td>
<td>0.230</td>
<td>0.468</td>
<td>0.073</td>
</tr>
<tr>
<td>100</td>
<td>0.274</td>
<td>0.164</td>
<td>0.366</td>
<td>0.057</td>
</tr>
<tr>
<td>125</td>
<td>0.233</td>
<td>0.168</td>
<td>0.294</td>
<td>0.044</td>
</tr>
<tr>
<td>150</td>
<td>0.206</td>
<td>0.161</td>
<td>0.283</td>
<td>0.037</td>
</tr>
<tr>
<td>175</td>
<td>0.164</td>
<td>0.121</td>
<td>0.203</td>
<td>0.030</td>
</tr>
<tr>
<td>200</td>
<td>0.170</td>
<td>0.106</td>
<td>0.230</td>
<td>0.038</td>
</tr>
<tr>
<td>250</td>
<td>0.136</td>
<td>0.099</td>
<td>0.172</td>
<td>0.026</td>
</tr>
</tbody>
</table>

2% lower (better) than the costs of solutions found when \(\hat{\beta} = \hat{\gamma} = 0\). However, such a small improvement in solution quality has a great impact in increasing computation time due to larger sample size \(N\).

Tables A5 and A6 summarize the values of the lower bounds, error gaps and running time for different values of sample size \(N\) and \(\hat{\beta}, \hat{\gamma}\). The error gap presented is smaller than 0.10%. This error gap is reduced as sample size \(N\) increases to 10,000 or 20,000, and for \(\hat{\beta} = \hat{\gamma} = 0.2\) and \(\hat{\beta} = \hat{\gamma} = 0.3\). Table A7 summarizes the best lower bound found for \(\beta = 0.2\) and 0.3 and \(\gamma = 0.2\) and 0.3. In these experiments, \(\hat{\beta} = \hat{\gamma} = 0.00\). The running time of the SAA Algorithm is much shorter as compared to solving the problems for \(\hat{\beta} = \beta\) and \(\hat{\gamma} = \gamma\) since the minimum sample size for which we can find a feasible solution, is smaller. The solutions found are within 2% error gap.

Based on the results of this analysis, we decided to set \(\hat{\beta} = \hat{\gamma} = 0.0\) and \(N = 125\) (for \(\beta = \gamma = 0.3\)), \(N = 250\) (for \(\beta = \gamma = 0.3\)) in our sensitivity analysis. Doing so, we get solutions of high quality in a reasonable amount of time.

A.4.4 Evaluating the Performance of SAA Algorithm to Solve \((Q)\):

To evaluate the effectiveness of the SAA in finding feasible solutions for \((Q)\), we vary the risk level \(\beta, \gamma\) and sample size \(N\). For each combination of \(\beta, \gamma\) and \(N\) we generate and solve 10 problems using the
Table A3: Solutions Returned by **SAA Algorithm** for \((P)\) \((\beta = \gamma = \hat{\beta} = \hat{\gamma} = 0.3)\)

<table>
<thead>
<tr>
<th>N</th>
<th>Avg</th>
<th>Min</th>
<th>Max</th>
<th>(\sigma)</th>
<th># of Feasible Solutions</th>
<th>Objective Value (in $1,000)</th>
<th>Average Run Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>0.317</td>
<td>0.307</td>
<td>0.327</td>
<td>0.006</td>
<td>0</td>
<td>-</td>
<td>45</td>
</tr>
<tr>
<td>2,000</td>
<td>0.308</td>
<td>0.299</td>
<td>0.318</td>
<td>0.007</td>
<td>2</td>
<td>28,991</td>
<td>8</td>
</tr>
<tr>
<td>3,000</td>
<td>0.305</td>
<td>0.298</td>
<td>0.312</td>
<td>0.005</td>
<td>2</td>
<td>28,992</td>
<td>7</td>
</tr>
<tr>
<td>5,000</td>
<td>0.306</td>
<td>0.302</td>
<td>0.314</td>
<td>0.004</td>
<td>0</td>
<td>-</td>
<td>600</td>
</tr>
<tr>
<td>7,500</td>
<td>0.304</td>
<td>0.301</td>
<td>0.308</td>
<td>0.002</td>
<td>0</td>
<td>-</td>
<td>1,265</td>
</tr>
<tr>
<td>10,000</td>
<td>0.303</td>
<td>0.299</td>
<td>0.308</td>
<td>0.003</td>
<td>2</td>
<td>28,988</td>
<td>3</td>
</tr>
<tr>
<td>20,000</td>
<td>0.302</td>
<td>0.298</td>
<td>0.304</td>
<td>0.002</td>
<td>2</td>
<td>28,986</td>
<td>2</td>
</tr>
</tbody>
</table>

Table A4: Solutions Returned by **SAA Algorithm** for \((P)\) \((\beta = \gamma = \hat{\beta} = \hat{\gamma} = 0.2)\)

<table>
<thead>
<tr>
<th>N</th>
<th>Avg</th>
<th>Min</th>
<th>Max</th>
<th>(\sigma)</th>
<th># of Feasible Solutions</th>
<th>Objective Value (in $1,000)</th>
<th>Average Run Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>0.217</td>
<td>0.210</td>
<td>0.225</td>
<td>0.006</td>
<td>0</td>
<td>-</td>
<td>44</td>
</tr>
<tr>
<td>2,000</td>
<td>0.209</td>
<td>0.200</td>
<td>0.219</td>
<td>0.007</td>
<td>0</td>
<td>-</td>
<td>130</td>
</tr>
<tr>
<td>3,000</td>
<td>0.206</td>
<td>0.202</td>
<td>0.213</td>
<td>0.004</td>
<td>0</td>
<td>-</td>
<td>243</td>
</tr>
<tr>
<td>5,000</td>
<td>0.206</td>
<td>0.203</td>
<td>0.209</td>
<td>0.002</td>
<td>0</td>
<td>-</td>
<td>607</td>
</tr>
<tr>
<td>7,500</td>
<td>0.203</td>
<td>0.201</td>
<td>0.206</td>
<td>0.002</td>
<td>0</td>
<td>-</td>
<td>1,440</td>
</tr>
<tr>
<td>10,000</td>
<td>0.204</td>
<td>0.200</td>
<td>0.208</td>
<td>0.002</td>
<td>1</td>
<td>29,147</td>
<td>2</td>
</tr>
<tr>
<td>20,000</td>
<td>0.202</td>
<td>0.200</td>
<td>0.205</td>
<td>0.002</td>
<td>1</td>
<td>29,152</td>
<td>8</td>
</tr>
</tbody>
</table>

Table A5: Lower Bounds for \((P)\) \((\beta = \gamma = \hat{\beta} = \hat{\gamma} = 0.3)\)

<table>
<thead>
<tr>
<th>N</th>
<th>2,000</th>
<th>3,000</th>
<th>10,000</th>
<th>20,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB Value</td>
<td>28,956,315</td>
<td>28,966,006</td>
<td>28,977,520</td>
<td>28,978,728</td>
</tr>
<tr>
<td>GAP (%)</td>
<td>0.10</td>
<td>0.07</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Average Run Time (sec.)</td>
<td>123</td>
<td>244</td>
<td>2,150</td>
<td>7,898</td>
</tr>
</tbody>
</table>

Table A6: Lower Bounds for \((P)\) \((\beta = \gamma = \hat{\beta} = \hat{\gamma} = 0.2)\)

<table>
<thead>
<tr>
<th>N</th>
<th>10,000</th>
<th>20,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB Value</td>
<td>29,137,519</td>
<td>29,140,580</td>
</tr>
<tr>
<td>GAP (%)</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Average Run Time (sec.)</td>
<td>2,251</td>
<td>8,439</td>
</tr>
</tbody>
</table>

Table A7: Lower Bounds for \((P)\)

<table>
<thead>
<tr>
<th>(\beta = \hat{\beta} = 0.0)</th>
<th>(\gamma = \hat{\gamma} = 0.0)</th>
<th>(\beta = \hat{\beta} = 0.2)</th>
<th>(\gamma = \hat{\gamma} = 0.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>125</td>
<td>250</td>
<td>28,978,728</td>
</tr>
<tr>
<td>GAP (%)</td>
<td>1.86</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Average Run Time (sec.)</td>
<td>0.20</td>
<td>0.36</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Table A8: Solutions returned by **SAA Algorithm** for $\beta = \gamma = 0.3$ ($\hat{\beta} = \hat{\gamma} = 0.0$)

<table>
<thead>
<tr>
<th>N</th>
<th>Ave</th>
<th>Min</th>
<th>Max</th>
<th>$\sigma$</th>
<th># of Feasible</th>
<th>Objective Value (in $1,000)</th>
<th>Avg Run Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.550</td>
<td>0.439</td>
<td>0.697</td>
<td>0.068</td>
<td>0</td>
<td>30.370</td>
<td>30.370</td>
</tr>
<tr>
<td>50</td>
<td>0.420</td>
<td>0.286</td>
<td>0.529</td>
<td>0.082</td>
<td>1</td>
<td>30.389</td>
<td>30.316</td>
</tr>
<tr>
<td>75</td>
<td>0.317</td>
<td>0.241</td>
<td>0.407</td>
<td>0.044</td>
<td>3</td>
<td>30.447</td>
<td>30.351</td>
</tr>
<tr>
<td>100</td>
<td>0.262</td>
<td>0.181</td>
<td>0.323</td>
<td>0.047</td>
<td>8</td>
<td>30.463</td>
<td>30.402</td>
</tr>
<tr>
<td>125</td>
<td>0.240</td>
<td>0.191</td>
<td>0.279</td>
<td>0.033</td>
<td>10</td>
<td>30.496</td>
<td>30.440</td>
</tr>
<tr>
<td>150</td>
<td>0.208</td>
<td>0.162</td>
<td>0.254</td>
<td>0.030</td>
<td>10</td>
<td>30.517</td>
<td>30.473</td>
</tr>
<tr>
<td>175</td>
<td>0.180</td>
<td>0.145</td>
<td>0.225</td>
<td>0.026</td>
<td>10</td>
<td>30.544</td>
<td>30.459</td>
</tr>
<tr>
<td>200</td>
<td>0.174</td>
<td>0.128</td>
<td>0.224</td>
<td>0.032</td>
<td>10</td>
<td>30.585</td>
<td>30.498</td>
</tr>
<tr>
<td>250</td>
<td>0.136</td>
<td>0.081</td>
<td>0.168</td>
<td>0.026</td>
<td>10</td>
<td>30.608</td>
<td>30.564</td>
</tr>
<tr>
<td>300</td>
<td>0.115</td>
<td>0.094</td>
<td>0.138</td>
<td>0.016</td>
<td>10</td>
<td>30.608</td>
<td>30.564</td>
</tr>
</tbody>
</table>

Based on the results of Tables A8, when $\hat{\beta} = \hat{\gamma} = 0.0$, we get 10 feasible solutions for $N \geq 125$. As $N$ increases, the number of constraints in $(\overline{Q})$ increases. The corresponding feasible region becomes smaller. This increases the likelihood that solutions found by solving $(\overline{Q})$ are feasible to $(Q)$ at this particular risk level. Additionally, increasing $N$ results in an increase of the cost of feasible solutions found. Based on the results in Table A9, when $\hat{\beta} = \hat{\gamma} = 0.0$, we get 10 feasible solutions for $N \geq 250$.

Table A9: Solutions returned by **SAA Algorithm** for $\beta = \gamma = 0.2$ ($\hat{\beta} = \hat{\gamma} = 0.0$)

<table>
<thead>
<tr>
<th>N</th>
<th>Ave</th>
<th>Min</th>
<th>Max</th>
<th>$\sigma$</th>
<th># of Feasible</th>
<th>Objective Value (in $1,000)</th>
<th>Avg Run Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.550</td>
<td>0.439</td>
<td>0.697</td>
<td>0.068</td>
<td>0</td>
<td>30.370</td>
<td>30.370</td>
</tr>
<tr>
<td>50</td>
<td>0.420</td>
<td>0.286</td>
<td>0.529</td>
<td>0.082</td>
<td>1</td>
<td>30.389</td>
<td>30.316</td>
</tr>
<tr>
<td>75</td>
<td>0.317</td>
<td>0.241</td>
<td>0.407</td>
<td>0.044</td>
<td>3</td>
<td>30.447</td>
<td>30.351</td>
</tr>
<tr>
<td>100</td>
<td>0.262</td>
<td>0.181</td>
<td>0.323</td>
<td>0.047</td>
<td>8</td>
<td>30.463</td>
<td>30.402</td>
</tr>
<tr>
<td>125</td>
<td>0.240</td>
<td>0.191</td>
<td>0.279</td>
<td>0.033</td>
<td>10</td>
<td>30.496</td>
<td>30.440</td>
</tr>
<tr>
<td>150</td>
<td>0.208</td>
<td>0.162</td>
<td>0.254</td>
<td>0.030</td>
<td>10</td>
<td>30.517</td>
<td>30.473</td>
</tr>
<tr>
<td>175</td>
<td>0.180</td>
<td>0.145</td>
<td>0.225</td>
<td>0.026</td>
<td>10</td>
<td>30.544</td>
<td>30.459</td>
</tr>
<tr>
<td>200</td>
<td>0.174</td>
<td>0.128</td>
<td>0.224</td>
<td>0.032</td>
<td>10</td>
<td>30.585</td>
<td>30.498</td>
</tr>
<tr>
<td>250</td>
<td>0.136</td>
<td>0.081</td>
<td>0.168</td>
<td>0.026</td>
<td>10</td>
<td>30.608</td>
<td>30.564</td>
</tr>
<tr>
<td>300</td>
<td>0.115</td>
<td>0.094</td>
<td>0.138</td>
<td>0.016</td>
<td>10</td>
<td>30.608</td>
<td>30.564</td>
</tr>
</tbody>
</table>

For this set of problems we did not develop lower bounds. This is mainly because, finding lower bounds required solving $(\overline{Q})$ for very large values of $N$. We conducted a few tests using $N = 10,000$ and $N = 20,000$, however, because of the large size of these problems, we could not find an optimal solution. As a result, we cannot comment on the quality of the solutions found by the SAA Algorithm for the decentralized model. However, we provide such an analysis for the centralized problem in Section A.4.3.
### A.5 Notation Used

#### SETS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B )</td>
<td>Biomass feedstocks</td>
</tr>
<tr>
<td>( I )</td>
<td>Suppliers</td>
</tr>
<tr>
<td>( P )</td>
<td>Unit farmgate costs</td>
</tr>
<tr>
<td>( \mathcal{X}, \mathcal{X}^L, \mathcal{X}^F )</td>
<td>Feasible region, feasible region of the leader’s problem, feasible region of ( i )-th follower’s problem</td>
</tr>
</tbody>
</table>

#### PARAMETERS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>Allowable ash content (in %)</td>
</tr>
<tr>
<td>( \tilde{\alpha}_{ib} )</td>
<td>Ash content of biomass ( b ) at supplier ( i ) (in %)</td>
</tr>
<tr>
<td>( \beta, \tilde{\beta}, \gamma, \tilde{\gamma} )</td>
<td>Risk parameters (in ( {0,1} ))</td>
</tr>
<tr>
<td>( c_p )</td>
<td>Unit farmgate cost (in $/ton)</td>
</tr>
<tr>
<td>( \bar{c} )</td>
<td>The unit cost of harvesting, collection, and storage of biomass ( b ) in price bracket ( p )</td>
</tr>
<tr>
<td>( c_{ibp} )</td>
<td>( c_p + t_{ib} + f_b )</td>
</tr>
<tr>
<td>( \bar{C}_b )</td>
<td>( c_p + t_{ib} + \bar{c}_{bp} )</td>
</tr>
<tr>
<td>( D_{ist} )</td>
<td>Distance (in km)</td>
</tr>
<tr>
<td>( \Delta() )</td>
<td>A measure of probability mass function value</td>
</tr>
<tr>
<td>( e_b )</td>
<td>Efficiency of the conversion process (in %)</td>
</tr>
<tr>
<td>( f_b )</td>
<td>Unit processing and inventory holding cost of a biorefinery (in $/ton)</td>
</tr>
<tr>
<td>( g_b )</td>
<td>Fixed unit transportation cost (in $/ton)</td>
</tr>
<tr>
<td>( \bar{h}_{ib} )</td>
<td>Thermal content of biomass feedstock ( b ) at supplier ( i ) (in %)</td>
</tr>
<tr>
<td>( k_{ibp}, \bar{k}_{ibp} )</td>
<td>Minimum/maximum amount of biomass available at price ( c_p ) (in tons)</td>
</tr>
<tr>
<td>( \lambda, \mu )</td>
<td>Penalty terms</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of scenarios</td>
</tr>
<tr>
<td>( p^* )</td>
<td>Optimal price bracket</td>
</tr>
<tr>
<td>( S_{ib} )</td>
<td>Amount of biomass feedstock ( b ) available at supplier ( i ) (in tons)</td>
</tr>
<tr>
<td>( \tau )</td>
<td>Annual thermal requirement (in BTU)</td>
</tr>
<tr>
<td>( t_{ib} )</td>
<td>Unit transportation cost (in $/ton)</td>
</tr>
<tr>
<td>( v_b )</td>
<td>Variable unit transportation cost (in $/\text{ton} \times \text{km})</td>
</tr>
<tr>
<td>( \vartheta )</td>
<td>Optimal objective function value</td>
</tr>
<tr>
<td>( z^*, c^l )</td>
<td>Optimal (quantity, price) decisions of the leader</td>
</tr>
</tbody>
</table>

#### DECISION VARIABLES

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_b )</td>
<td>The door price of biomass ( b )</td>
</tr>
<tr>
<td>( X_{ib} )</td>
<td>Amount of biomass ( b ) shipped from supplier ( i ) (in tons)</td>
</tr>
<tr>
<td>( X_{ibp} )</td>
<td>Amount of biomass ( b ) shipped from supplier ( i ) at price ( p ) (in tons)</td>
</tr>
<tr>
<td>( Z_{ibp} )</td>
<td>Binary variable which takes the value 1 when the amount of biomass ( b ) available at ( i ) falls in price bracket ( p ), and takes the value 0 otherwise</td>
</tr>
<tr>
<td>( Z_p )</td>
<td>Binary variable which takes the value 1 when the amount of biomass purchased falls in price bracket ( p ), and takes the value 0 otherwise</td>
</tr>
<tr>
<td>( V, W, U, J )</td>
<td>Slack variables</td>
</tr>
</tbody>
</table>
Appendix B  
Integrated Biorefinery Operations

B.1 Detailed Notations

<table>
<thead>
<tr>
<th>SETS:</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
</tr>
<tr>
<td>$T$</td>
</tr>
<tr>
<td>$E^g$</td>
</tr>
<tr>
<td>$E^{mill}$</td>
</tr>
<tr>
<td>$E^p$</td>
</tr>
<tr>
<td>$E^r$</td>
</tr>
<tr>
<td>$E$</td>
</tr>
<tr>
<td>$N$</td>
</tr>
<tr>
<td>$\delta_i$</td>
</tr>
<tr>
<td>$S$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARAMETERS:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
</tr>
<tr>
<td>$h$</td>
</tr>
<tr>
<td>$f(t)$</td>
</tr>
<tr>
<td>$\kappa_t$</td>
</tr>
<tr>
<td>$\bar{v}_i(\kappa_t)$</td>
</tr>
<tr>
<td>$\bar{i}_i(\kappa_t)$</td>
</tr>
<tr>
<td>$\bar{r}$</td>
</tr>
<tr>
<td>$S$</td>
</tr>
<tr>
<td>$\epsilon$</td>
</tr>
<tr>
<td>$i_{s1}$</td>
</tr>
<tr>
<td>$i_{s2}$</td>
</tr>
<tr>
<td>$i_r$</td>
</tr>
<tr>
<td>$\gamma_0$</td>
</tr>
<tr>
<td>$\gamma_i$</td>
</tr>
<tr>
<td>$\phi_i(\kappa_t)$</td>
</tr>
<tr>
<td>$\bar{l}_i$</td>
</tr>
<tr>
<td>$c_i$</td>
</tr>
<tr>
<td>$\bar{c}_i$</td>
</tr>
<tr>
<td>$i^m$</td>
</tr>
<tr>
<td>$\bar{x}_i(\kappa_t)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RANDOM VARIABLES (SCENARIO-DEPENDENT PARAMETERS):</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{0ts}$</td>
</tr>
<tr>
<td>$m_{its}$</td>
</tr>
<tr>
<td>$d_{its}$</td>
</tr>
<tr>
<td>$\rho_{jts}$</td>
</tr>
<tr>
<td>$\theta_{its}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DECISION VARIABLES:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{it}$</td>
</tr>
<tr>
<td>$I_{i0s}$</td>
</tr>
<tr>
<td>$I_{its}$</td>
</tr>
<tr>
<td>$X_{0ts}$</td>
</tr>
<tr>
<td>$X_{its}$</td>
</tr>
</tbody>
</table>

Table A10: Detailed Mathematical Notation
B.2 Detailed Formulation

Introducing a surrogate linear objective function. The true objective of this problem is to minimize the total operational cost per dry tons of biomass processed during the planning horizon, which is computed using unit “dollar per dry ton” ($/dt). In this case the energy consumption and operation cost function, $\hat{f}(\cdot)$ is as follows:

$$\hat{f}(X_{ts}, I_{ts}, \omega_s) = \sum_{i \in N} \frac{T(e_i + c_i)}{\sum_{t \in T} (1 - m_{i,ts})X_{i,ts}}.$$  

As a result the true objective that we evaluate is given by:

$$\min \sum_{i \in E} h_i I_{i0} + \frac{1}{S} \left[ \sum_{i \in N} T(e_i + c_i) \left( \sum_{s=1}^{S} \sum_{t \in T} (1 - m_{i,ts})X_{i,ts} \right) \right].$$

However, this results in a non-convex objective function, which makes the problem hard to solve. Notice that, in order to minimize the total cost per dry ton, we need to maximize the amount of biomass processed, while using minimum initial inventory. By this observation, we consider the following surrogate linear objective function, which makes the problem much easier to solve.

$$\min \sum_{i \in E} h_i I_{i0} - \left( \frac{1}{S} \left[ \sum_{s=1}^{S} \sum_{t \in T} (1 - m_{i,ts})X_{i,ts} \right] \right).$$

Note that two terms (i.e., holding cost and average amount of biomass processed) of this surrogate objective function have different units. In order to resolve this we used a weighted sum of the two terms, where the weight of the initial holding cost is much smaller.

Use of this surrogate objective function leads to the following chance-constrained stochastic linear

108
program:

\[
(P) := \min \sum_{i \in E} h_i I_{i0} - \left( \frac{1}{S} \sum_{s=1}^{S} \sum_{j \in T} (1 - m_{i,s}) X_{i,s} \right)
\]

s.t. 0 \leq V_{it} \leq \bar{\phi}_i(\kappa_t),  \quad \forall i \in N, t \in T,

0 \leq I_{i0} \leq \bar{\phi}_i(\kappa_0),  \quad \forall i \in E^m, 

X_{0ts} = \gamma_0 \delta_{0ts} V_{1t}, \quad \forall t \in T, s \in S

X_{its} = (1 - \varphi_i(\kappa_i)) \sum_{j \in \delta^-_i} X_{jts}, \quad \forall i \in E^v, t \in T, s \in S

X_{its} \leq \gamma_i d_{its} V_{it}, \quad \forall i \in E^v, t \in T, s \in S

X_{its} = \sum_{j \in \delta^-_i} X_{jts}, \quad \forall i \in E^v \setminus \{i_{b1}, (i_{b2})\}, t \in T, s \in S

X_{i_{b2}, ts} = (1 - \theta_{ts}) X_{i_{b1} ts}, \quad \forall t \in T, s \in S

X_{(i_{b2}), ts} = \theta_{ts} X_{i_{b1} ts}, \quad \forall t \in T, s \in S

X_{its} = \gamma_i d_{its} V_{it}, \quad \forall i \in E^m, t \in T, s \in S

I_{ts} = I_{i(t-1)s} + \sum_{j \in \delta^-_i} X_{jts} - X_{its}, \quad \forall i \in E^m, t \in T \setminus \{1\}, s \in S

I_{its} \geq \frac{1}{t} d_{its}, \quad \forall i \in E^m, t \in T, s \in S

I_{its} \leq \bar{\phi}_i d_{its}, \quad \forall i \in E^m, t \in T, s \in S

X_{its} = (1 - \varphi_i(\kappa_i)) \left( 1 - \frac{\bar{\phi}_i}{t} \right) \sum_{j \in \delta^-_i} X_{jts} + I_{i(t-1)s}, \quad \forall i \in E^{mll}, t \in T, s \in S

I_{its} = I_{i(t-1)s} + \sum_{j \in \delta^-_i} X_{jts} - X_{its}, \quad \forall i \in E^{mll}, t \in T, s \in S

I_{its} \leq \bar{\phi}_i d_{its}, \quad \forall i \in E^{mll}, t \in T, s \in S

(1 - m_{i,s}) X_{its} \leq \bar{\phi}_i(\kappa_i), \quad \forall i \in E^p, t \in T, s \in S

(1 - m_{i,s}) X_{i,s} \leq \bar{\phi}_i(\kappa_i), \quad \forall t \in T, s \in S

X_{0ts}, X_{its} \geq 0, \quad \forall i \in N, t \in T, s \in S

I_{i0} \geq 0, \quad \forall i \in E^m \cup E^{mll},

I_{its} \geq 0, \quad \forall i \in E^m \cup E^{mll}, s \in S

\frac{1}{S} \sum_{s=1}^{S} \frac{1}{\bar{\phi}} \sum_{t \in T} (1 - m_{i,t}) X_{i,t} \geq r \quad \geq 1 - \bar{\phi}_i.
The bypass ratio in the separation unit is calculated as follows:

\[
\theta_{ts} := \begin{cases} 
\max\{0.5 - 0.4(\rho_{50}^{\text{ts}} - 6.35)/(\rho_{50}^{\text{ts}} - \rho_{10}^{\text{ts}}), 0\} & \text{for } \rho_{50}^{\text{ts}} \geq 6.35 \\
\min\{0.5 + 0.4(6.35 - \rho_{50}^{\text{ts}})/(\rho_{90}^{\text{ts}} - \rho_{50}^{\text{ts}}), 1\} & \text{for } \rho_{50}^{\text{ts}} < 6.35,
\end{cases}
\]  

(28)

where, the value 6.35 (in mm) corresponds to the screen size used in the separation unit.

We explain the constraints in formulation (\(\hat{P}\)) as follows:

- Constraints (6) and (7) correspond to the constraints (3.3) and constraints (3.7) in the succinct formulation (\(\hat{P}\)) in the main document. They represent the bounds on the equipment processing speed and initial inventory, respectively.

- Constraints (8), (9), (14), and (19) represent the flow calculations for processing and storage equipment. These constraints correspond to constraints (3.10) in the succinct formulation (\(\hat{P}\)) in the main document. For example, constraints (9) calculate the biomass flow with respect to the moisture and dry matter losses during the grinding process.

- Constraints (10) represent the upper limit on the amount of biomass flow from each transportation equipment \(i \in E\). These constraints correspond to constraints (3.11) in the succinct formulation (\(\hat{P}\)) in the main document.

- In the transportation equipment, the biomass flowing into the equipment equals to the biomass flowing from it. Constraints (11), (12), and (13) represent these flow balance equations, just like constraints (3.9) in the succinct formulation (\(\hat{P}\)) in the main document.

- Constraints (15) and (16) are the inventory balance constraints of the storage equipment \(i \in E^m\). They are also a part of constraints (3.10) in the succinct formulation (\(\hat{P}\)) in the main document.

- Constraints (17) and (18) set the upper and lower thresholds for the inventory level. The inventory upper bound comes from the storage equipment’s capacity. The lower threshold is required for consistent flow out of the storage equipment \(E^m\). Although the volume of a storage equipment is fixed, variations in the biomass density impact the amount of biomass allowed to be stored. These constraints are the detailed version of constraints (3.12) in the succinct formulation (\(\hat{P}\)) in the main document.
Pelleting process takes $t_i$ units of time (which is less than the chosen time period $t$) in the pelleting mill $i$. The pelleting mill has an in-process storage capacity, which keeps the biomass during the pelleting process. Constraints (20) and (21) represent the inventory balance and inventory capacity in pelleting mill $E^{mill}$, respectively. In the succinct formulation in the main documents, we included these constraints within constraints (3.10) and (3.12), respectively.

Processing equipment and the reactor have limits on the amount of biomass flowing into them. Constraints (22) and (23) represent these upper limits. These constraints are part of constraints (3.10) in the succinct formulation ($\hat{P}$) in the main document.

### B.3 Robust Optimization

The real-world problems often include uncertain data. There are several sources of data uncertainty: i) the data might not be available at the time of problem-solving, such as future demand ii) exact value of the data might not be measured (i.e., the real value fluctuates around the recorded measurement), such biomass moisture content, iii) there might be implementations errors that limit the capability to implement the decided solution exactly, this can be incorporated in technology parameters as data uncertainty. Let’s define a generic linear programming (LP) problem with uncertain data:

$$\min_x \{c^T x : Ax \leq b\}_{(c,A,b) \in U}$$

where $U$ is the given uncertainty set for the data $(c, A, b)$.

Ben-Tal et al. (2009) identifies the following three implicit assumptions of robust optimization:

A.1. The decision vector $x$ represents "here-and-now" decisions: the specific numeric values of the components of $x$ must be identified (by the solution) before the realization of the data is observed.

A.2. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within the prespecified uncertainty set $U$.

A.3. The constraints of the uncertain LP in question are “hard” (i.e., the decision maker cannot tolerate violations of constraints when the data is in $U$).

These assumptions results in identifying “immunized against uncertainty” solutions to the uncertain LP problem Ben-Tal et al. (2009). Here “immunized against uncertainty” means that a robust feasible solution
will be feasible to the original problem regardless of the realization of the data, as long as it is within $U$. The robust counterpart (RC) of the uncertain LP is the following
\[
\min_x \left\{ \sup_{c, A, b \in U} c^\top x : Ax \leq b, \forall (c, A, b) \in U \right\}.
\]

If we analyze the RC model, we observe that constraints must be satisfied for all realizations of the uncertainty within $U$, and in the objective the function $\sup_{c, A, b \in U}$ ensures that the solution is assessed by the worst-case scenario with respect to $(c, A, b)$.

Constraints (3.5) and (3.6) represent the upper and lower bounds for the amount of biomass flowing from each equipment and the amount of inventory stored in the equipment, respectively. In constraints (3.5) and (3.6), we have the decisions $V_t$, $I_0$, $I_t$, and $X_t$. It is clear that $V_t$ and $I_0$ are “here-and-now” decisions. Additionally, $I_t$ and $X_t$ depends on “here-and-now” decisions and the uncertain data $\omega$. Thus, constraints (3.5) and (3.6) can be reformulated such that they only depend on “here-and-now” decisions. This means we need to make sure they are feasible before the realization of the data is observed. Bounds on the biomass flow and inventory level are physical limitations of the equipment. Thus, in our model, we need to satisfy these limitations regardless of the realizations of $\omega$. These features of constraints (3.5) and (3.6) show that they have the assumptions A.1. and A.3. implicitly.

Sample average approximation is not sufficient to ensure the feasibility of these constraints for all realizations of $\omega$ Chen and Luedtke (2021). Due to this we replaced (3.5) and (3.6) with their robust counterparts. We start with a reformulation and replace $X_t(\omega)$ and $I_t(\omega)$ with their representation based on $V_t$ and $\omega$. As a result, constraints (3.5) and (3.6) are replaced by the following constraints:
\[
d'(\omega, V_t) \leq b_t, \forall t \in T, \omega \in \Omega, \quad (29)
\]
here $b_t$ represent the component of vector $b$ associated with time $t \in T$. The robust counterpart of (29) can be formulated as follows:
\[
\sup_{\omega \in \Omega} d'(\omega, V_t) \leq b_t, \forall t \in T \quad (30)
\]
Here, $\sup_{\omega \in \Omega} d'(\omega, V_t)$ is an optimization problem by itself with respect to $\omega$. We call this the inner optimization problem. To solve the inner optimization, we perform the following four main steps:

(i) We constructed a box uncertainty set for each random variable in $\omega$. For example for the random
variable \( \omega_j \) (i.e., the random variable in the \( j^{th} \) dimension of \( \omega \)) we constructed the uncertainty set of \([\omega_j, \overline{\omega}_j]\). In here \( \omega_j, \overline{\omega}_j \) represent the lower and upper bounds of the random variable \( \omega_j \).

\[
\tilde{d}_{0t} \in [d_{0t}, \overline{d}_{0t}]
\]
\[
\tilde{m}_{it} \in [m_{it}, \overline{m}_{it}]
\]
\[
\tilde{d}_{it} \in [d_{it}, \overline{d}_{it}]
\]

Box uncertainty sets provide a probabilistic guarantee that the robust solution will be feasible to the original uncertain constraint regardless of the \( \omega \) realization. However, box uncertainty can be too pessimistic when used in practice Gorissen et al. (2015). Different uncertainty sets are proposed in literature to address this pessimism. Such as ellipsoidal uncertainty set, let’s call it \( \Omega^E \), in Ben-Tal et al. (2009).

\[
\Omega^E = \{\omega : ||\omega||_2 \leq \Gamma\},
\]

where \( \Gamma = \sqrt{2 \ln(1/\epsilon)} \) is the safety parameter of the ellipsoid. \( \Omega^E \) ensures that the robust optimal solution satisfied the original uncertainty constraint with probability at least \( 1 - \epsilon \).

(ii) Random variables in \( \omega \) are independent of each other with respect to \( t \). Thus constraints (30) can be handled separately with respect to \( t \).

\[
\sup_{\omega \in \Omega} d'(\omega, V_1) \leq b_1
\]
\[
\sup_{\omega \in \Omega} d'(\omega, V_2) \leq b_2
\]
\[
\cdots
\]

(iii) \( d'(\omega, V_1) \) is not convex with respect to \( \omega \) due to multilinear terms (e.g., \( \omega_i \times \omega_j \times \omega_k \), where \( i, j, \) and \( k \) represent a dimension of \( \omega \)). We can conservatively approximate these multilinear terms with a new independent random variable \( \xi \). For example let the uncertainty sets of \( \omega_i, \omega_j, \omega_k \) be \([0, 2]\), \([3, 5]\), and \([2, 20]\), respectively. Then the multiplication \( \omega_i \times \omega_j \times \omega_k \) can take values between 0 and 200. Thus, we constructed the uncertainty for the random variable \( \hat{\xi} := \omega_i \times \omega_j \times \omega_k \) as \([0, 200]\).

We had the following multilinear terms presented in the left hand side of each row in our model. We
replaced them with the new random variable represented on the right hand sides.

\[
\begin{align*}
\tilde{d}_t \tilde{\theta}_t & \rightarrow \xi^1_t \\
\tilde{d}_t \tilde{m}_t & \rightarrow \xi^2_t \\
(1 - \tilde{\theta}_t)\tilde{d}_t (1 - \tilde{m}_t) & \rightarrow \xi^3_t \\
\tilde{m}_t \tilde{d}_t & \rightarrow \xi^4_t \\
\tilde{d}_t (1 - \tilde{m}_{12t}) & \rightarrow \xi^5_t \\
\tilde{d}_{t-1}(1 - \tilde{m}_{12t}) & \rightarrow \xi^6_t
\end{align*}
\]

For these newly introduced random variables $\xi^j_t$, we constructed uncertainty sets considering the range of each random variable and the value of their multiplications as explained in the example.

\[
\xi^j_t \in \left[\xi^j_t, \bar{\xi}^j_t\right], \ j \in \{1, 2, \ldots, 6\}.
\]

(iv) After the conservative approximation is applied, constraints (30) are all linear with respect to $\xi$. Thus we can utilize LP duality to solve the inner optimization.

After applying the steps (i)-(iii), we have the following constraints:

\[
\max_{\xi \in \Xi} D\xi V_t \leq b_t, \ \forall t \in T.
\] (31)

Let $l$ be the dimension of $\xi$ and $\Xi := \bigcup_{j=1}^{l} \left[\underline{\omega}_j, \bar{\omega}_j\right]$. At this point we do row by row robustification of constraint (31).

\[
\max_{\xi \in \Xi} \sum_{j \in \{1, \ldots, l\}} (D_{r,j}\xi_j V_t) \leq b_{r,t}, \ \forall r \in \{1, \ldots, m\}, t \in T, \quad (32)
\]
where $m$ is the number of rows. Using the LP duality we reformulate these constraints as:

\[
\sum_{j \in \{1, \ldots, l\}} (\xi_j \pi_j^1 - \xi_j \pi_j^2) \leq b_{r,t}, \ \forall r \in \{1, \ldots, m\}, t \in T, \quad (33)
\]
\[
\pi_j^1 - \pi_j^2 \leq D_{r,j} V_{it}, \ \forall j \in \{1, \ldots, l\}, r \in \{1, \ldots, m\}, t \in T, \quad (34)
\]

where $\pi_j^1, \pi_j^2$ are dual variables associated with lower and upper bound of $\xi_j$.

Furthermore, we observe that in most of the rows random variable $\xi_j$ is multiplied with a single decision variable (e.g., $V_{it}, i \in N$) in (32). Thus, we can easily identify if the worst case is associated with the upper or lower bound of $\xi_j$. For these rows, we can directly plug in the associated worst case. For example, consider the following constraint

\[
\sum_{j \in \{1, \ldots, l\}} (D_{r,j} \xi_j) V_{it} \leq b_{r,t}, \ \forall t \in T,
\]

where we only have the decision variable $V_{it}$. Also, consider that the worst case scenario comes from the upper bound of the random variables $\{\xi_j\}_{j \in \{1, \ldots, l\}}$. Then, we can directly use the following constraint

\[
\sum_{j \in \{1, \ldots, l\}} (D_{r,j} \xi_j^U) V_{it} \leq b_{r,t}, \ \forall t \in T.
\]

**Extensive Robustified Constraints:**

The reformulation explained in (29) results in constraints that are heavy in notation. For the ease of following, we decided to replaced multiplications of some problem parameters with succinct representations.

\[
\hat{\gamma}_{0t} = \gamma_0 (1 - \phi_1 - \varphi_{1t}(\kappa_t)), \ \forall t \in T,
\]
\[
\bar{\gamma}_{0t} = \gamma_0 (1 - \phi_1 - \varphi_{1t}(\kappa_t))(1 - \phi_5 - \varphi_{5t}(\kappa_t)), \ \forall t \in T,
\]
\[
\gamma_{12t} = (1 - \varphi_{12t}(\kappa_t)), \ \forall t \in T,
\]
\[
\tau = \frac{\hat{t}_{12}}{t}.
\]

As a result of the four steps explain above we end up with the several deterministic constraints. These are the robust counterparts of the constraints (6)-(23). Constraints (6)-(23) will be replaced by the
following constraints in the robust model.

\[
\begin{align*}
\gamma_0 d_{0t} V_{1t} - \gamma_0 \xi_{1t} V_{1t} - \gamma_2 d_{1t} V_{2t} & \leq 0, & \forall t \in T, \\
\gamma_0 d_{0t} V_{1t} - \gamma_0 \xi_{2t} V_{1t} & \leq \bar{x}_1(\kappa_t), & \forall t \in T, \\
\gamma_0 d_{0t} V_{1t} - \gamma_0 \xi_{1t} V_{1t} - \gamma_3 d_{1t} V_{3t} & \leq 0, & \forall t \in T, \\
\gamma_0 \xi_{1t} V_{1t} - \gamma_1 d_{1t} V_{11t} & \leq 0, & \forall t \in T, \\
\gamma_0 \xi_{1t} V_{1t} & \leq \bar{x}_5(\kappa_t), & \forall t \in T, \\
\gamma_0 d_{0t} V_{1t} - \gamma_0 \xi_{1t} V_{1t} - \gamma_6 d_{6t} V_{6t} & \leq 0, & \forall t \in T, \\
\gamma_0 d_{0t} V_{1t} - \gamma_0 \xi_{1t} V_{1t} - \gamma_7 d_{7t} V_{7t} & \leq 0, & \forall t \in T, \\
\gamma_0 d_{0t} V_{1t} - \gamma_0 \xi_{1t} V_{1t} - \gamma_8 d_{8t} V_{8t} & \leq 0, & \forall t \in T, \\
I_{9,0} + \sum_{j=1}^{t} ((\gamma_{0j} \xi_{j} + \gamma_{0j} d_{0j} + \gamma_{0j} \xi_{1j}) V_{1j} - \gamma_9 d_{9j} V_{9j}) & \geq \tilde{\xi}_9 d_{9t}, \forall t \in T, \\
I_{9,0} + \sum_{j=1}^{t} ((\gamma_{0j} \xi_{j} + \gamma_{0j} d_{0j} + \gamma_{0j} \xi_{1j}) V_{1j} - \gamma_9 d_{9j} V_{9j}) & \leq \bar{\xi}_9 d_{9t}, \forall t \in T, \\
\tilde{d}_{0t} \pi_{9t} - d_{9t} \pi_{9t} & \leq 0, & \forall t \in T, \\
\pi_{9t}^1 - \pi_{9t}^2 = \gamma_9 V_{9t} - \gamma_{10} V_{10t}, & \forall t \in T, \\
\gamma_{12t} \gamma_9 d_{9t} V_{9t} & \leq \bar{\iota}_{12t} d_{12t}, & \forall t \in T, \\
\gamma_{12t}(1 - \tau) \gamma_9 d_{9t} V_{9t} + \gamma_{12t} \tau \gamma_9 d_{9(t-1)} V_{9(t-1)} & \leq \gamma_{13} d_{12t} V_{13t}, \forall t \in T, \\
\gamma_9 d_{9t} V_{9t} - \gamma_9 \xi_{4t} V_{9t} & \leq \bar{x}_{12}(\kappa_t), & \forall t \in T, \\
\gamma_{12,1} \gamma_9 \xi_{14} V_{9,1} & \leq r, & \forall t \in T, \\
\gamma_{12t}(1 - \tau) \gamma_9 \xi_{14} V_{9t} + \gamma_{12t} \tau \gamma_9 \xi_{14} V_{9(t-1)} & \leq r, & \forall t \in T.
\end{align*}
\]
In this section, we share the data tables for the biorefinery operations study.

Table A11: Energy Consumption and Fixed Costs

<table>
<thead>
<tr>
<th>Equipment</th>
<th>High moisture</th>
<th>Med moisture</th>
<th>Low moisture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Energy Cost</td>
<td>Fixed Cost</td>
<td>Energy Cost</td>
</tr>
<tr>
<td></td>
<td>($/hr)</td>
<td>($/hr)</td>
<td>($/hr)</td>
</tr>
<tr>
<td>Bale conveyor</td>
<td>0.12</td>
<td>0.48</td>
<td>0.12</td>
</tr>
<tr>
<td>Grinder 1</td>
<td>1.72</td>
<td>31.32</td>
<td>1.47</td>
</tr>
<tr>
<td>Screw conveyor-6</td>
<td>0.29</td>
<td>11.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Drag chain conveyor-5</td>
<td>0.17</td>
<td>1.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Drag chain conveyor-6</td>
<td>0.29</td>
<td>1.56</td>
<td>0.29</td>
</tr>
<tr>
<td>Screw conveyor-1</td>
<td>0.17</td>
<td>1.20</td>
<td>0.17</td>
</tr>
<tr>
<td>Grinder 2</td>
<td>3.33</td>
<td>13.85</td>
<td>1.11</td>
</tr>
<tr>
<td>Screw conveyor-2</td>
<td>1.15</td>
<td>4.22</td>
<td>1.15</td>
</tr>
<tr>
<td>Screw conveyor-4</td>
<td>0.29</td>
<td>11.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Metering bin</td>
<td>0.63</td>
<td>9.98</td>
<td>0.63</td>
</tr>
<tr>
<td>Screw conveyor-5</td>
<td>0.29</td>
<td>6.14</td>
<td>0.29</td>
</tr>
<tr>
<td>Drag chain conveyor-1</td>
<td>0.12</td>
<td>1.61</td>
<td>0.12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>14.63</strong></td>
<td><strong>105.30</strong></td>
<td><strong>10.16</strong></td>
</tr>
<tr>
<td></td>
<td><strong>8.67</strong></td>
<td><strong>104.98</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table A12: Particle Size Distribution Percentiles

<table>
<thead>
<tr>
<th>Moisture Level</th>
<th>Biomass Processed in Grinder 1</th>
<th>Biomass Processed in Grinder 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho^{50}$</td>
<td>$\rho^{90}/\rho^{10}$</td>
</tr>
<tr>
<td>Low</td>
<td>[1.90 – 2.00]</td>
<td>[11.5 – 13.5]</td>
</tr>
<tr>
<td>Medium</td>
<td>[2.30 – 2.40]</td>
<td>[11.0 – 13.0]</td>
</tr>
<tr>
<td>High</td>
<td>[1.70 – 1.80]</td>
<td>[9.0 – 11.00]</td>
</tr>
</tbody>
</table>

Table A13: Moisture and Dry Matter Changes

<table>
<thead>
<tr>
<th>Moisture Level</th>
<th>Moisture Loss</th>
<th>Dry Matter Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grinder 1 (%)</td>
<td>Grinder 2 (%)</td>
</tr>
<tr>
<td>Low</td>
<td>0.50</td>
<td>0.70</td>
</tr>
<tr>
<td>Medium</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>High</td>
<td>4.77</td>
<td>4.00</td>
</tr>
</tbody>
</table>

Table A14: Equipment Infeed Rate Limits

<table>
<thead>
<tr>
<th>Moisture Level</th>
<th>Low Moisture (dt/hr)</th>
<th>Med Moisture (dt/hr)</th>
<th>High Moisture (dt/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grinder 1</td>
<td>Grinder 2</td>
<td>Pelleting Mill</td>
</tr>
<tr>
<td>Low</td>
<td>5.23</td>
<td>4.53</td>
<td>2.20</td>
</tr>
<tr>
<td>Medium</td>
<td>5.23</td>
<td>2.80</td>
<td>1.59</td>
</tr>
<tr>
<td>High</td>
<td>4.76</td>
<td>3.81</td>
<td>3.34</td>
</tr>
</tbody>
</table>
B.5 Data Tables - Adaptive Biomass Processing Operations

Tables in this section presents the changes we considered from in the adaptive biomass processing operations compared to Appendix B.4.

Table A15: Particle size distribution percentiles.

<table>
<thead>
<tr>
<th>Moisture Level</th>
<th>Biomass Processed in Grinder 1</th>
<th>Biomass Processed in Grinder 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho_{50}$ (mm)</td>
<td>$\rho^{90}/\rho^{10}$</td>
</tr>
<tr>
<td>Low</td>
<td>1.95</td>
<td>12.5</td>
</tr>
<tr>
<td>Medium</td>
<td>2.35</td>
<td>12.0</td>
</tr>
<tr>
<td>High</td>
<td>1.75</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Table A16: Secondary grinder bypass ratios.

<table>
<thead>
<tr>
<th>Moisture Level</th>
<th>Bypass Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Bypass Ratio</td>
<td>85.7</td>
</tr>
</tbody>
</table>

B.6 Regression Analysis for Biomass Density

In this section, we summarize the statistical results of the regression analyses.

Table A17: Regression Analysis Statistics

<table>
<thead>
<tr>
<th>Regression</th>
<th>$R^2$</th>
<th>$\alpha_i^0$</th>
<th>$m_{it}$</th>
<th>$\rho_{50}^{50}$</th>
<th>$\rho_{50}^{90}/\rho_{50}^{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(16)</td>
<td>0.956</td>
<td>$4.1 \times 10^{-142}$</td>
<td>$9.5 \times 10^{-134}$</td>
<td>$1.3 \times 10^{-69}$</td>
<td>$6.9 \times 10^{-1}$</td>
</tr>
<tr>
<td>(17)</td>
<td>0.945</td>
<td>$2.1 \times 10^{-69}$</td>
<td>$2.4 \times 10^{-75}$</td>
<td>$2.5 \times 10^{-39}$</td>
<td>$7.3 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
B.7 A Bisection Search Based Heuristic Algorithm for Solving the SAA Problem

(\hat{P})

Algorithm 3 A bisection search based algorithm for solving the SAA problem (\hat{P})

\textbf{Input:} Set \( \pi = 10^7, \xi = 0, \delta = 0.01 \) and \( \sigma = 0.01 \)

1: Set \( \pi \leftarrow \pi \).
2: Solve model (\hat{P}) and get an optimal solution \( \hat{U}_s \).
3: Set a counter \( C \leftarrow 0 \).
4: for \( s \in S \) do
5: if \( \hat{U}_s > 0 \) then
6: \( C \leftarrow C + 1 \)
7: end if
8: end for
9: if \( C \geq \hat{\epsilon}N + \sigma \) then
10: Status \leftarrow \text{"INFEASIBLE: Target R is not achievable"}.
11: else
12: while \( \frac{\pi}{2} < \delta \) do
13: \( \pi \leftarrow \pi \leftarrow \frac{\pi}{2} \)
14: Solve model (\hat{P}) and get an optimal solution \( \hat{U}_s \).
15: Set a counter \( C \leftarrow 0 \).
16: for \( s \in S \) do
17: if \( \hat{U}_s > 0 \) then
18: \( C \leftarrow C + 1 \)
19: end if
20: end for
21: if \( C \geq \hat{\epsilon}N + \sigma \) then
22: \( \pi \leftarrow \frac{\pi}{2} \)
23: else
24: \( \pi \leftarrow \frac{\pi}{2} \)
25: end if
26: end while
27: Let \( \hat{V}, \hat{X}, \hat{I} \) be the optimal solution of model (\hat{P}) in the final iteration.
28: Return \( \hat{V}, \hat{X}, \hat{I} \)
29: end if
Bibliography


