Integrating Biomass Quality Variability in Stochastic Supply Chain Modeling and Optimization for Large-scale Biofuel Production

Krystel K. Castillo-Villar[a,*], Sandra Eksioglu[b], Milad Taherkhorsandi[a]

[a] *Department of Mechanical Engineering, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA.*
[b] *Department of Industrial Engineering, Clemson University, Freeman Hall, Clemson, SC 29634 USA.*

Abstract

The production of biofuels using second-generation feedstocks has been recognized as an important alternative source of sustainable energy and its demand is expected to increase due to regulations such as the Renewable Fuel Standard. However, the pathway to biofuel industry maturity faces unique, unaddressed challenges.

This paper presents an optimization model which quantifies and controls the impact of biomass quality variability on supply chain related decisions and technology selection. We propose a two-stage stochastic programming model and associated efficient solution procedures for solving large-scale problems to (1) better represent the random nature of the biomass quality (defined by moisture and ash contents) in supply chain modeling, and (2) assess the impact of these uncertainties on the supply chain design and planning.

The proposed model is then applied to a case study in the state of Tennessee. Results show that high moisture and ash contents negatively impact the unit delivery cost since poor biomass quality requires the addition of quality control activities. Experimental results indicate that supply chain cost could increase as much as 27% to 31% when biomass quality is poor. We assess the impact of the

*Corresponding author
Email address: Krystel.Castillo@utsa.edu (Krystel K. Castillo-Villar)

biomass quality on the topological supply chain. Our case study indicates that biomass quality impacts supply chain costs; thus, it is important to consider the impact of biomass quality in supply chain design and management decisions.
Keywords: Quality Costing; Biomass; Bioenergy; Biofuels; Stochastic Programming; L-shaped; Optimization; Supply Chain Network Design.

---

1. Introduction

In recent years, industry has seen the advent of highly complex, large-scale supply chains (SCs), which have become increasingly difficult to analyze and optimize when using conventional modeling techniques and solution procedures. The design of large-scale SCs is a crucial task in a plethora of scientific fields, including advanced manufacturing and electric power networks, among others. A relevant example is found in the *emerging bioenergy industry*. This industry requires *sophisticated mathematical models and solution approaches* to enhance biomass SCs by *integrating biomass quality control principles; and biomass quality uncertainties in the SC design and management decision making process.*

Biofuel has been recognized as an alternative source of renewable energy [46]. Its demand and production is expected to increase in the upcoming years [4, 25, 50], primarily due to the legislation enacted by the United States of America Energy Independence and Security Act of 2007. An outcome of this act is the renewable fuel standards (RFS) [16]. The billion ton study lead by the Oak Ridge National Laboratory indicates that the country can sustainably produce over a billion tons of biomass (i.e., forest biomass/residues, agricultural biomass/residues and energy crops) annually [13]. However, the delivery of biomass required to meet the goals set by the RFS is particularly challenging. This is mainly because of the physical properties of biomass, which is bulky and widely geographically dispersed. First-generation biomass such as corn and soybean, among others have higher energy density, lower ash content, and lower collections/transportation costs as compared to agricultural and forest waste. However, these types of biomass raised the national debate of food versus fuel, which considers the use of marginal lands for energy crop production and the land and water usage changes in the bioeconomy [34, 12]. This controversy was one of the main reasons for the development of second-generation biofuels that use energy crops, agricultural and forest waste (not suitable for neither human nor livestock consumption). However, second generation biomass exhibits more biomass quality variability (e.g., higher ash and moisture contents) than first generation biomass. This paper focuses on the use of energy crops (i.e., switchgrass) to produce second-generation biofuels.

The pathway to industry maturity faces two main challenges. One of the challenges is developing technologies which ensure a cost efficient conversion of biomass to biofuels and are robust to biomass quality variations. The second challenge is developing cost efficient biofuel supply chains which are robust to variabilities in biomass supply and costs. This paper contributes to the area of biomass supply chain design and management by developing a stochastic programming model which captures the impacts of biomass quality on supply chain related decisions. The goal is to quantify and control the impacts of biomass quality variability on supply chain related costs and technology selection.

Although bioenergy is an emerging industry, the biomass supply system inherited models and underlying assumptions from the well-established agricultural and logging industries. Therefore, the objective of most (if not all) of the biomass feedstock logistics models is to reduce the overall costs, under the assumption that the biomass quality specifications and process requirements are similar to forage and pulpwood [28]. The single objective of minimizing the total (of purchasing, logistics and processing) costs may have a considerable negative impact on the expected profit and on the performance of bioenergy SCs because, in practice, these bioenergy supply systems work with highly variable and/or poor quality biomass, which cause important economic losses. A recent report from Idaho National Laboratory [28] raised the concern that research on feedstock quality is still lacking and that the traditional models disregard quality-related issues by driving down the logistics cost. This emphasis of cost over quality is exemplified by the current pricing structure for biomass, which is based on the measure "dollar per dry ton" instead of "dollar per *clean* dry carbohydrate." Practitioners who have scaled up to pilot-scale operations (which require large quantities of feedstock) have experienced considerable differences between "pristine" and "field-run" biomass [24].

Moreover, scale-up risk becomes a very important parameter to consider in the bioenergy industry as new technologies evolve from the laboratory to commercial settings [3]. For example, consider the undesirable scenario where the biorefinery equipment, designed to work with a biomass moisture content of approximately 10%, has to work with biomass with moisture content of 30% in a given year. In addition, consider the financial losses if one load of feedstock yields 90 gallons/ton and another load yields 60 gallons/ton. [28] demonstrate via a number of case studies that these scenarios are very likely to occur in practice. The biomass quality is dependent upon the moisture, ash, sugar contents, and particle morphology, among others. Thus, ignoring biomass quality variations and the associated costs when modeling biomass SCs may yield costly results that will only be discovered after the operations at a biorefinery have begun. The development of quality control methods for biomass is a largely unmet topic in the literature. The model presented in this paper enables us to quantify the impacts of biomass quality and technology uncertainties in the performance of the SC.

In order to address these challenges, we propose *(1) a two-stage stochastic programming model which integrates biomass supply and quality variabilities in supply chain modeling; and (2) a solution approach which is used to solve large-scale problems and evaluate the impacts of biomass quality on supply chain decisions.* The proposed model and solution procedures contribute to *(1) better represent the random nature of the biomass quality in SC modeling, and (2) assess the impact of these uncertainties on the SC design and planning.* Uncertainty and risk are two of the main challenges faced by enterprises with complex SCs. A proper assessment of the uncertainties and risks related to supply availability and quality, and opportunity costs when making long-term decisions is vital for the profitability and sustainability of this enterprise. The inception of decision support tools based on stochastic programming models and quality control principles is a new concept in the field of SC. Indeed, most (if not all) of the literature on biofuel planning tools is focused on deterministic models [14, 23, 52, 41]. Only a few stochastic programming models exist in this literature [15]. Generally speaking, stochastic models are underutilized in the SC field due to their complexity; that is, these techniques typically result in models that require an enormous computational effort for solving large-scale practical instances. For these reasons, the primary analytic tools currently used in the bioenergy industry tend to be fairly ad hoc.

This paper is organized as follows. In section 2, we present a discussion of the related literature that positions our work in context. Section 3 provides a description of the problem addressed and the associated novel mathematical model. Section 4 presents the solution approach based on extensions of the L-shaped and multicut L-shaped algorithms to solve large-scale problems. Section 5 presents a case study that uses realistic data from Tennessee. Section 6 shows a computational study that evaluates the algorithmic performance of the proposed solution procedures and discusses managerial insights. Finally, in Section 7, we provide concluding remarks as well as future research lines.

2. Literature Review

The majority of the mathematical models in the bioenergy SC field are modeled as Mixed Integer Linear Programming (MILP) or Linear Programming (LP) optimization. In general, there is a lack of approaches that include uncertainty and risk modeling for bioenergy logistics [15]. For example, [10] developed a linear programming model which is used as a planning tool for the assessment of the costs associated with biomass transferred from producers located in close proximity to a centrally located plant. The objective was to minimize the transportation costs and the capacity expansion costs at storage sites for individual producers. The results estimated biomass delivery costs to biorefineries and also recommended the shipping and capacity expansion schedules for each producer. [11] addressed the scheduling of both single and multiple feedstocks in a single digester system for biogass to methane production systems. They solved the multiple feedstock problem using a decomposition approach which separates the problem into one master problem and a number of subproblems. The master problem allocates time at the digester for each feedstock. The subproblems schedule batches within these time allocations. To maximize biogas production, the available times, biomass quantities, biogas production rates and storage decay rates were considered during the planning horizon. [30] proposed a general optimization model involving the selection of fuel conversion technologies, capacities, biomass locations, and the logistics of transportation. They used GAMS to implement the MILP model. The results depicted the overall profits and supply network designs of the system and also illustrated the parameters having major effects on the overall economics.

Mixed integer stochastic programming has been employed for modeling and optimizing bioenergy supply chain systems that involve uncertainty. [29] studied a supply chain network model which is focused in the southeastern region of the United States. This model identifies biomass supply locations, facility sittings, capacities for two kinds of fuel conversion processing, and the logistics transportation. The model proposed is a two-stage stochastic program, with the first stage decisions identifying the size and location of the preprocessing plants; and the second stage decisions identifying the product flow by scenario. The objective is the maximization of the expected profit over the different scenarios. A global sensitivity analysis using Monte Carlo simulation was also performed in order to estimate the performance of the system as some problem parameters change. [9] utilized a mixed integer stochastic programming model to provide the strategic planning of bioenergy supply chain systems and optimal feedstock resource allocation in an uncertain decision environment. They developed a case study using data from California; and solved the problem via a Lagrangean relaxation-based decomposition algorithm (i.e., progressive hedging method or horizontal decomposition).

The L-shaped algorithm is one of the solution approaches for stochastic optimization problems which has been used extensively in the literature [31]. Both, the L-shaped and multicut L-shaped methods, decompose the master problem into as many subproblems as the total number of scenarios. These problems are solved iteratively until a stopping criteria is met. Other approaches to solve stochastic optimization models use variations of the branch-and-bound algorithm [2, 42, 45, 17] or, when appropriate, use extensions of Benders decomposition algorithm [8, 7, 22, 43, 45]. In the bioenergy supply chain field, [19] presented a bicriterion, multiperiod, stochastic mixed-integer linear programming model for the optimal design of hydrocarbon biorefinery supply chains under supply and demand uncertainties. To minimize the expected annualized cost and the financial risk simultaneously, they proposed a model which captures multiple conversion technologies, feedstock seasonality and fluctuation, geographical diversity, demand variation, government incentives, biomass degradation, and risk management. They propose a multi-cut L-shaped algorithm to reduce the computational time when solving large-scale instances. Four case studies of hydrocarbon biorefinery supply chain in the State of Illinois were solved using the proposed algorithm. [36] presented a two-stage stochastic programming model to design and manage biodiesel supply chains. They proposed an L-shaped algorithm. They used a Lagrangian relaxation algorithm to solve the master problem since it is an integer program. These authors developed a case study using data from Mississippi. The model optimizes both costs and emissions in the supply chain. The results elucidated the impact of carbon regulatory mechanisms on supply chain costs and emissions and also the effectiveness of the stochastic programming.

3. An Integrated Product Quality and Supply Chain Design Model

3.1. Modeling Product Quality

Two product characteristics, which are indicators of biomass quality, are moisture and ash contents. Let the biomass moisture content be a random variable, $\epsilon(t)$, which depends on a specified mean value t. We assume that $\epsilon(t)$ follows a triangular distribution in $[at, bt]$ with a probability density function as presented below. This assumption is based on experimental results conducted with real switchgrass under different storage and harvesting conditions [51].

$$f_{\epsilon(t)}(e) = \begin{cases} \frac{2(e-at)}{(bt-at)(t-at)} & at \leq e \leq t \\ \frac{2(bt-e)}{(bt-at)(bt-t)} & t < e \leq bt \\ 0 & \text{o/w} \end{cases} \tag{1}$$

Similarly, we assume that ash content is a random variable $\vartheta(\delta)$, whose distribution is a function of the mean value, δ. We model ash content using a triangular distribution for $\vartheta(\delta)$ in $[c\delta, d\delta]$.

Processes currently used to produce biofuels do have a number of requirements with respect to biomass quality. For example, processes that rely on the thermochemical conversion technology have a targeted value of moisture content of no more than 10%. We will refer to the technology target as t_k. When this constraint is violated, a failure cost equal to $\$q$ per unit is incurred. This is the cost of mechanically drying biomass to reduce its moisture to acceptable levels. The expected cost for not meeting the quality requirements is computed as the square of the deviation between the value of the quality characteristic and the target value. This cost can be regarded as the opportunity cost and is expressed as:

$$M(\epsilon(t)) = my_1^2, \tag{2}$$

where, $y_1 = max(\epsilon(t) - t_k, 0)$. The expected quality loss is given by:

$$\phi_1(\epsilon(t)) = \int_{-\infty}^{+\infty} M(\epsilon(t)) f_{\epsilon(t)}(e) de.$$

Similarly, processes which use a thermochemical conversion technology, rely on using biomass with no more than 1% of ash content. Let δ_k represent the ash content targeted by technology k. Thus, the opportunity cost for not meeting the ash specification is:

$$N(\vartheta(\delta)) = ny_2^2, \tag{3}$$

where, $y_2 = max(\vartheta(\delta) - \delta_k, 0)$ and, the expected quality loss is given by:

$$\phi_2(\vartheta(\delta)) = \int_{-\infty}^{+\infty} N(\vartheta(\delta)) f_{\vartheta(\delta)}(v) dv.$$

3.2. A Two-Stage Stochastic Model

Stochastic programming models assume that the probability distributions governing the data are known or can be estimated [6]. The most extensively studied stochastic programming models are the two-stage (linear) models. These models capture the timing of decisions in the SC, where, the first-stage decisions are made right now and without full knowledge of future events. These future events have random outcomes. These random outcomes and the first-stage decision impact the future (second-stage) decisions in the supply chain. In such a model, a recourse decision is made in the second-stage to account for any non-beneficial effects that might have resulted from the first-stage actions. The optimal policy corresponds to a single first-stage decision and a set of recourse decisions (for each random outcome) that define which second-stage action should be taken [44]. Two-stage stochastic programming models typically assume that the random event can be described using discrete random variables with known probability distributions.

The two-stage stochastic location-transportation model identifies facility locations that minimize the total of location and expected transportation costs. The model we propose in this paper is one of its many extensions. The following sets are defined, I is the set of suppliers, J is the set of potential biorefinery locations and K is the set of biomass conversion technologies. Let Z_{jk} ($\forall j \in J$, $k \in K$) be the first stage decision variables, which take the value 1 if a facility that uses technology k is located in j, and take the value 0 otherwise. Let

$X_{ijk} \in \mathcal{X} \subseteq R_+^{|I|\times|J|\times|K|}$ be the second stage decision variables which represent the amount of biomass delivered from supplier i to facility j which uses technology k. Let l_{jk} be the equivalent annualized investment cost for opening a biorefinery in location j using technology k. Biomass quality at supplier i is not constant, but it rather fluctuates from one season to the next, and from one year to the next. Let ω be a random variable which represents biomass quality. Its probability density function is $f_\omega(.)$. The following stochastic programming model minimizes the total of location and expected transportation costs in the supply chain.

$$Minimize : TC(Z) = \sum_{j\in J}\sum_{k\in K} l_{jk}Z_{jk} + \mathcal{Q}(Z)$$

Subject to: (P)

$$Z_{jk} \in \{0,1\} \qquad \forall j \in J; k \in K.$$

Where,

$$\mathcal{Q}(Z) = E_\omega Q(Z, X, \omega) = \int_{-\infty}^{+\infty} Q(Z, X, o) f_\omega(o) d(o)$$
$$X \in \mathcal{X}.$$

3.3. Integrating Product Quality in the Supply Chain Model

The model presented in this section takes an integrated view of key variables that impact supply chain design and management decisions of biofuel plants, such as, location, transportation, technology selection, and product quality. The goal of this model is to minimize the total supply chain costs by capturing the trade-offs that exist between location and transportation costs; technology selection and quality costs; facility location and quality costs.

Biomass quality impacts differently plants that use different technologies. Biomass quality requirements are different in a plant that uses a thermochemical conversion process, versus a plant that uses a biochemical conversion process.

Even within thermochemical conversion, biomass quality requirements may vary based on the particular process used, such as, pyrolysis, gasification or combustion. Thus, with each technology $k \in K$, we associate two parameters: t_k, which represents the requirements with respect to moisture content; and δ_k, which represents the requirements with respect to ash content. Each supplier provides a product that has specific moisture and ash contents. Based on historical data, suppliers commit to deliver biomass with $t_i\%$ moisture content and $\delta_i\%$ ash content. These values, $t_i\%$ and $\delta_i\%$, represent the expected moisture level and ash content of biomass supplied by supplier i. Indeed, the moisture and ash contents are random variables that follow a triangular distribution (see Section 3.1).

Let ω be a discrete variable with density function $f_\omega(o) = P[\omega = o] = P(o)$ for $o \in \Omega$. Consider the special case of the problem with 2 scenarios ($|\Omega| = 2$). Scenario 1 assumes that weather conditions are rather dry in the region under study, and scenario 2 assumes a rainy weather. Therefore, under scenario 1, moisture content is low. Low moisture content implies low quality costs since the amount of energy required to dry biomass could potentially be zero. Moisture content also impacts the amount of biomass available at a supplier. Typically, dry weather negatively impacts the productivity of agricultural products, and thus, the amount of agricultural waste available. Under scenario 2 moisture content is high.

Under scenario 1, the random variable e_i – which represents moisture content of supplier i – lies on the lower side of the triangular distribution; thus, $at_i \leq e_i \leq t_i$. Let $f^1_{\epsilon(t_i)}(e_i) = \frac{2(e_i - at_i)}{(bt_i - at_i)(t_i - at_i)}$ be the moisture content density function of supplier i under scenario 1. If t_k is the target moisture level under technology k; the expected quality loss for scenario 1 assuming that supplier i is selected, would be:

$$\phi_1(\epsilon(t_i)) = \int_{at_i}^{t_i} \frac{2m(e_i - t_k)^2(e_i - at_i)}{(bt_i - at_i)(t_i - at_i)} de_i =$$

$$\frac{2m}{(bt_i - at_i)(t_i - at_i)} \int_{at_i}^{t_i} (e_i - t_k)^2 (e_i - at_i) de_i =$$

$$m\left(\gamma_i' t_k^2 - \beta_i' t_k + \alpha_i'\right). \tag{4}$$

where γ_i', β_i' and α_i' are analytical expressions derived in the Appendix A.

To generalize, the expected quality loss due to moisture content under scenario o and for a given t_k is equal to:

$$c_i'(t_k, o) = m\left(\gamma_i'(o) t_k^2 - \beta_i'(o) t_k + \alpha_i'(o)\right). \tag{5}$$

Similarly, each supplier commits to providing biomass which has a particular ash content by following harvesting techniques which result in biomass with $\delta_i\%$ ash content. This is indeed the expected ash content based on historical data. However, the actual ash content is a random variable, ϑ_i, which follows a triangular distribution as described in section 3.1. The expected quality loss from shipments from supplier i under scenario o and for a given δ_k is a fixed constant equal to:

$$c_i(\delta_k, o) = n\left(\gamma_i(o)\delta_k^2 - \beta_i(o)\delta_k + \alpha_i(o)\right). \tag{6}$$

The derivations of $\gamma_i(o)$, $\beta_i(o)$ and $\alpha_i(o)$ are shown in the Appendix A.

The following two-stage, stochastic model is proposed to minimize the total of location, transportation, technology selection, and quality costs in the SC and it is referred as Model (Q). In model (Q), we use a few additional problem parameters, such as, $s_i(o)$, which represents the amount of biomass available at supplier i under scenario o. Let g_{ik} represent a conversion factor (in gallons/ton) of biomass to biofuel. The value of this factor depends on the type of biomass supplied from i and technology adapted at facility k. Let l_{jk} denote the facility location costs; ν_{jk} denote the production capacity of facility j when adapting technology k; d denote the total demand for bioenergy; and c_{ij} denote the unit delivery cost from supplier i to facility j.

$$Minimize: TC(Z, X, \omega) = \sum_{k\in K}\sum_{j\in J} l_{jk} Z_{jk} +$$

$$\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}\sum_{o\in\Omega} P(o)\left[c_{ij} + c_i'(t_k, o) + c_i(\delta_k, o)\right] X_{ijk}(o)$$

Subject to: (Q)

$$\sum_{j\in J}\sum_{k\in K} X_{ijk}(o) \leq s_i(o) \qquad \forall i \in I, o \in \Omega \qquad (7)$$

$$\sum_{i\in I} g_{ik} X_{ijk}(o) \leq \nu_{jk} Z_{jk} \qquad \forall j \in J, k \in K, o \in \Omega \qquad (8)$$

$$\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} g_{ik} X_{ijk}(o) \geq d \qquad \forall o \in \Omega \qquad (9)$$

$$\sum_{k\in K} Z_{jk} \leq 1 \qquad \forall j \in J \qquad (10)$$

$$X_{ijk}(o) \in R^+ \qquad \forall i \in I, j \in J, k \in K, o \in \Omega \qquad (11)$$

$$Z_{jk} \in \{0, 1\} \qquad \forall j \in J, k \in K. \qquad (12)$$

In model (Q), constraints (7) give an upper bound on the amount of biomass available at supplier i under scenario o. Constraints (8) connect the continuous flow variables X_{ijk} with the binary variables Z_{jk}. These constraints restrict biofuel production to the maximum capacity of the biorefinery. Constraints (9) enforce total biofuel demand satisfaction. Constraints (10) limit the selection of one technology per facility. Constraints (11) are the non-negativity constraints, and (12) are the binary constraints.

4. A Solution Approach for the Integrated Model

Model (Q) is a two-stage SP model where the first-stage decision variables are integers and the second-stage decisions are continuous. We propose an L-shaped and a multicut L-shaped method to solve this stochastic optimization problem.

4.1. L-shaped method

Consider the following equivalent formulation of problem (Q).

$$Minimize : TC(Z) = \sum_{k\in K}\sum_{j\in J} l_{jk} Z_{jk} + \mathcal{G}$$

Subject to:

$$\sum_{k\in K} Z_{jk} \leq 1 \qquad \forall j \in J \tag{13}$$

$$\mathcal{G} \geq \mathcal{Q}(Z) \tag{14}$$

$$Z_{jk} \in \{0,1\} \qquad \forall j \in J, k \in K. \tag{15}$$

where $\mathcal{Q}(Z) = E_o(Q(Z,o))$, and

$$Q(Z,o) = \min_X : \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} \bar{c}_{ijk}(o)X_{ijk}(o) + p\pi(o)$$

Subject to:

$$\sum_{j\in J}\sum_{k\in K} X_{ijk}(o) \leq s_i(o) \qquad \forall i \in I \tag{16}$$

$$\sum_{i\in I} g_{ik}X_{ijk}(o) \leq \nu_{jk}Z_{jk} \qquad \forall j \in J, k \in K \tag{17}$$

$$\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} g_{ik}X_{ijk}(o) + \pi(o) = d \tag{18}$$

$$X_{ijk}(o) \in R^+ \qquad \forall i \in I, j \in J, k \in K \tag{19}$$

where, $\bar{c}_{ijk}(o) = c_{ij} + c'_i(t_k, o) + c_i(\delta_k, o)$. The literature refers to above formulation of (Q) as the *Master* problem, and $Q(Z,o)$ as the *Subproblems* (SP(o)) (for $o \in \Omega$) [49]. The Master problem is an Integer Program (IP). To solve an IP Master problem Lagrangian relaxation or valid inequalities which improve the solution quality and reduce the required computational time are suitable approaches. The Subproblems (SP(o)) are Linear Programs (LP), thus, relatively easy to solve. We solve the Subproblems efficiently using a problem specific algorithm. Next, we describe a few extensions of the L-shaped algorithm with the goal of improving its performance.

Note that, in Subproblems (SP(o)) uncertainty only affects the right-hand-side of constraints (16) (i.e., biomass supply). The recourse matrix characterized by left-hand-sides in equations (16)-(18) and the transfer matrix characterized by right-hand side of equations (17) are independent of randomness. Therefore, the above two-stage stochastic program has a *Fixed Recourse* [32].

Let $\mathcal{N}$ represents the set of solutions Z_{jk} ($\forall j \in J, k \in K$) which satisfy constraints (13) to (15). Let $\overline{Z}_{jk}^{n}$ represent the $n-th$ solution in this set. Subproblems (SP(o)) are feasible even if $Z_{jk} = 0$ ($\forall j \in J, k \in K$). In this case, $\pi(o) = d$. $\pi(o)$ represent the demand unmet via this supply chain. To discourage such solutions, we penalize the unmet demand via a high penalty cost p in the objective function. Thus, Subproblems (SP(o)) are always *Feasible*. This is why the proposed Bender's decomposition algorithm does not generate feasibility cuts.

Let β_o, α_{io} and λ_{jko} be the dual variables of the primal (SP(o)) Subproblem. The following is the corresponding dual formulation for given $\overline{Z}_{jk}^{n}$. Note that, $\overline{Z}_{jk}^{n}$ appear only in the objective function of the dual formulation. Thus, as we update the values of $\overline{Z}_{jk}^{n}$ (for $n = 1, 2, ...$) the optimal solution may iterate among vertexes of the same feasible region.

$$D(\overline{Z}^{n}, o) = \max_{\alpha,\beta,\lambda} : d\beta_o - \sum_{i \in I} s_i(o)\alpha_{io} - \sum_{j \in J}\sum_{k \in K} \nu_{jk}\overline{Z}_{jk}^{n}\lambda_{jko}$$

Subject to:

$$\begin{aligned}
g_{ik}\beta_o - \alpha_{io} - g_{ik}\lambda_{jk} &\leq \bar{c}_{ijk}(o) && \forall i \in I, j \in J, k \in K && (20)\\
\beta_o &\leq p && && (21)\\
\alpha_{io} &\in R^{+} && \forall i \in I && (22)\\
\lambda_{jk} &\in R^{+} && \forall j \in J, k \in K. && (23)
\end{aligned}$$

Let $(\alpha_{io}^{n}, \beta_{o}^{n}, \lambda_{jko}^{n})$ denote the optimal solution to $D(\overline{Z}^{n}, o)$ for fixed values of $\overline{Z}_{jk}^{n}$. By duality, the following holds true:

$$Q(\overline{Z}^{n}, o) = \left(d\beta_o^{n} - \sum_{i \in I} s_i(o)\alpha_{io}^{n} - \sum_{j \in J}\sum_{k \in K} \nu_{jk}\lambda_{jko}^{n}\overline{Z}_{jk}^{n} \right)$$

By convexity of $Q(Z, o)$, the following is also true:

$$Q(Z, o) \geq \left(d\beta_o^{n} - \sum_{i \in I} s_i(o)\alpha_{io}^{n} - \sum_{j \in J}\sum_{k \in K} \nu_{jk}\lambda_{jko}^{n}Z_{jk} \right)$$

We now take the expectations of these two functions to obtain the following relationships.

$$\mathcal{Q}(Z) = \sum_{o \in \Omega} (P(o)Q(Z,o)) \geq \sum_{o \in \Omega} P(o) \left(d\beta_o^n - \sum_{i \in I} s_i(o)\alpha_{io}^n - \sum_{j \in J} \sum_{k \in K} \nu_{jk} \lambda_{jko}^n Z_{jk} \right)$$

We use this relationship to develop the following equivalent formulation of (Q) which we refer to as (EQ).

$$Minimize: TC(Z) = \sum_{k \in K} \sum_{j \in J} l_{jk} Z_{jk} + \mathcal{G}$$

Subject to:

$$\sum_{k \in K} Z_{jk} \leq 1 \qquad \forall j \in J \tag{24}$$

$$\sum_{o \in \Omega} P(o) \left(d\beta_o^n - \sum_{i \in I} s_i(o)\alpha_{io}^n - \sum_{j \in J} \sum_{k \in K} \nu_{jk} \lambda_{jko}^n Z_{jk} \right) \leq \mathcal{G} \quad n \in \mathcal{N} \tag{25}$$

$$Z_{jk} \in \{0,1\} \qquad \forall j \in J, k \in K. \tag{26}$$

In this formulation, the number of constraints (25) is governed by the size of set $\mathcal{N}$, which could be a large number. Recall that $\mathcal{N}$ represents the set of solutions Z_{jk} ($\forall j \in J, k \in K$) which satisfy constraints (13) to (15). These constraints limit the number of non-zero Z_{jk} variables (in each iteration of the Bender's algorithm) to at most $|J|$. Thus, the size of $\mathcal{N}$ is equal to 1 plus the number of subsets of set J, $2^{|J|}$. For this reason, instead of solving (EQ), we solve the following reduced model formulation which we refer to as (REQ). This model is solved iteratively, and $1 \leq l \leq |\mathcal{N}|$ represents the iteration number.

$$Minimize: TC(Z) = \sum_{k \in K} \sum_{j \in J} l_{jk} Z_{jk} + \mathcal{G}$$

Subject to:

$$\sum_{k \in K} Z_{jk} \leq 1 \qquad \forall j \in J \tag{27}$$

$$\sum_{o\in\Omega} P(o)\left(d\beta_o^n - \sum_{i\in I} s_i(o)\alpha_{io}^n - \sum_{j\in J}\sum_{k\in K}\nu_{jk}\lambda_{jko}^n Z_{jk}\right) \leq \mathcal{G} \quad n = 1,\ldots,l \quad (28)$$

$$Z_{jk} \in \{0,1\} \qquad \forall j \in J, k \in K. \quad (29)$$

The L-shaped algorithm is presented in Table 1. *Step 1* of the algorithm solves (REQ) which is a relaxation of the Master problem and therefore its objective function value serves as a lower bound for (Q). In *Step 2* each subproblem (SP(o)) is solved provided solutions Z^n from Step 1. An upper bound is calculated using the solutions from the Master problem and the Subproblems. In the first iteration of this algorithm, since $n = 0$, the optimal solution to (EQ) is $Z_{jk} = 0$. This implies that no facilities are open in the supply chain. In this iteration, the solution to the Subproblems are $\pi(o) = d$. In *Step 3* we check the relative gap between the best bounds generated so far. If the gap is less than a threshold ϵ the algorithm terminates. Otherwise, an optimality cut is added to the Master problem, and the problem is resolved.

4.1.1. Trust region cuts:

Based on [39], cutting plane-based algorithms (such as, Benders decomposition) exhibit unstable behavior in their initial iterations. That means, solutions tend to oscillate from one feasible region to another which leads to slow convergence. Therefore, [40] suggests the use the following trust region inequalities which bound the Hamming distance [20] of the solutions found in consecutive iterations of the algorithm.

Let $\overline{Z}_{jk}^n$ (for $j \in J, k \in K$) be the solution obtained from solving (REQ) during iteration n. Let $\mathcal{Z}^{n+} = \{(j,k)|\overline{Z}_{jk}^n = 1, \forall j \in J, k \in K\}$. The following inequality is added to (REQ).

$$\sum_{(j,k)\in\mathcal{Z}^{n+}} (1 - Z_{jk}^{n+1}) + \sum_{(j,k)\notin\mathcal{Z}^{n+}} Z_{jk}^{n+1} \leq 1. \quad (30)$$

These inequalities force the solutions generated during iterations n and $n+1$ of the algorithm to differ by at most one variable. These inequalities expedite

Table 1: An L-shaped algorithm for problem (Q)

STEP 0:
  Initialize: ε, $n \leftarrow 1$, $LB \leftarrow -\infty, UB \leftarrow +\infty$
STEP 1:
  Solve (REQ) to obtain Z^n_{jk}, $TC(Z^n)$
  If $(TC(Z^n) > LB)$ **Then**
    $LB \leftarrow TC(Z^n)$
  End If
STEP 2:
  For all $o \in \Omega$ **Do**
    Solve (SP(o)) to obtain $\alpha^n_{io}, \beta^n_o, \lambda^n_{jk}$
  End for
  If $TC(Z^n) + \sum_{o\in\Omega} \big(P(o)D(\overline{Z}^n, o)\big) - \mathcal{G} < UB$ **Then**
    $UB \leftarrow TC(Z^n) + \sum_{o\in\Omega} \big(P(o)D(\overline{Z}^n, o)\big) - \mathcal{G}$
  End If
STEP 3:
  If $((UB - LB)/UB < \varepsilon)$ **Then STOP**
  Else
    Add to (REQ):
    $\sum_{o\in\Omega} P(o) \Big(d\beta^n_o - \sum_{i\in I} s_i(o)\alpha^n_{io} - \sum_{j\in J}\sum_{k\in K} \nu_{jk}\lambda^n_{jko}Z_{jk}\Big) \leq \mathcal{G}$
    $n \leftarrow n + 1$; **GoTo STEP 1**
  End If

the running time of the algorithm during its initial iterations. Later on, we drop these constraints in order to maintain the feasibility of (REQ).

4.1.2. Multi-cut L-shaped algorithm:

The multi-cut L-shaped algorithm was introduced by [5] to enhance the convergence of the L-shaped algorithm. In each iteration of the L-shaped algorithm (Table 1), one single optimality cut (constraint (28)) is added to formulation (REQ). Instead, in each iteration of the multi-cut L-shaped algorithm one could add as many as $|\Omega|$ cuts, one cut per scenario. In this case, formulation (REQ) is slightly modified to the following. We call this formulation (Ω-REQ)

$$Minimize : TC(Z) = \sum_{k\in K}\sum_{j\in J} l_{jk}Z_{jk} + \sum_{o\in\Omega} P(o)\mathcal{G}(o)$$

Subject to: (27), (29)

$$d\beta^n_o - \sum_{i\in I} s_i(o)\alpha^n_{io} - \sum_{j\in J}\sum_{k\in K} \nu_{jk}\lambda^n_{jko}Z_{jk} \quad \leq \quad \mathcal{G}(o) \quad n = 1, \ldots, l, o \in \Omega \tag{31}$$

To solve this problem we slightly adjust the algorithm presented in Table 1. The modification made is the following: STEP 1 solves the modified (Ω-REQ) instead of (REQ).

4.1.3. Solving the subproblems:

The structure of problems (SP(o)) resembles the transportation problem. It is well known that the solutions to the transportation problem can be degenerate. Therefore, the dual of this problem can have multiple optimal solutions. This fact leads to the possibility of multiple optimality cuts being generated in each iteration of the algorithm. The literature indicates that, if one is to pick amongst these cuts the "strongest" one, then, the approach would result in improvements of the existing algorithm [48]. In the context of our problem, let $\sum_{o\in\Omega} P(o)\left(d\beta_o^{n_1} - \sum_{i\in I} s_i(o)\alpha_{io}^{n_1} - \sum_{j\in J}\sum_{k\in K}\nu_{jk}\lambda_{jko}^{n_1}Z_{jk}\right) \leq \mathcal{G}$ and $\sum_{o\in\Omega} P(o)\left(d\beta_o^{n_2} - \sum_{i\in I} s_i(o)\alpha_{io}^{n_2} - \sum_{j\in J}\sum_{k\in K}\nu_{jk}\lambda_{jko}^{n_2}Z_{jk}\right) \leq \mathcal{G}$ be two cuts generated during the n-th iteration of the L-shaped algorithm, each corresponding to one of the two optimal solutions to the dual of (SP(o)). The first cut is stronger than the second if the following holds true for any $Z_{jk} \in \mathcal{N}$.

$$\sum_{o\in\Omega} P(o)\left(d\beta_o^{n_1} - \sum_{i\in I} s_i(o)\alpha_{io}^{n_1} - \sum_{j\in J}\sum_{k\in K}\nu_{jk}\lambda_{jko}^{n_1}Z_{jk}\right) \geq$$

$$\sum_{o\in\Omega} P(o)\left(d\beta_o^{n_2} - \sum_{i\in I} s_i(o)\alpha_{io}^{n_2} - \sum_{j\in J}\sum_{k\in K}\nu_{jk}\lambda_{jko}^{n_2}Z_{jk}\right)$$

In order to generate tight cuts we follow a similar procedure as the two-phase approach proposed by [38] for the capacitated facility location problem. In the dual formulation of (SP(o)), the dual values associated with $\overline{Z}_{jk}^n$ parameters which are zero, do not have any impact on the optimal objective function value. Recall that, at most $|J|$ (from a total of $|J| * |K|$) $\overline{Z}_{jk}^n$ parameters are non-zero. Therefore, when $\overline{Z}_{jk}^n = 0$, we can modify its coefficient $-\nu_{jk}\lambda_{jko}$ without changing the objective function value while, of course, maintaining the feasibility of the corresponding solution (satisfying constraints (32)). The following is a summary of the two-phase process.

Phase 1: We solve a reduced model formulation of the dual of (SP(o)) to obtain a set of values for the dual variables λ^n_{jko} associated with the variables $\overline{Z}^n_{jk} > 0$, as well as, to obtain the optimal α^n_{io} $(\forall i \in I)$, and β^n_o. Let $J^> = \{j \in J | \overline{Z}^n_{jk} > 0\}$ and $K^> = \{k \in K | \overline{Z}^n_{jk} > 0\}$. Formulation (R1-SP(0)) presented below is solved to find these optimal solutions.

$$D^>(\overline{Z}^n, o) = \max_{\alpha,\beta,\lambda} : d\beta_o - \sum_{i \in I} s_i(o)\alpha_{io} - \sum_{j \in J^>} \sum_{k \in K^>} \nu_{jk} \overline{Z}^n_{jk} \lambda_{jko}$$

Subject to:

$$\begin{aligned}
g_{ik}\beta_o - \alpha_{io} - g_{ik}\lambda_{jk} &\leq \bar{c}_{ijk}(o) & \forall i \in I, j \in J^>, k \in K^> \quad (32)\\
\beta_o &\leq p & (33)\\
\alpha_{io} &\in R^+ & \forall i \in I \quad (34)\\
\lambda_{jk} &\in R^+ & \forall j \in J^>, k \in K^>. \quad (35)
\end{aligned}$$

Phase 2: We fix the values of α^n_{io} to $\overline{\alpha}^n_{io}$ $(\forall i \in I)$, and β^n_o to $\overline{\beta}^n_o$ as determined in the first phase of this procedure. Next, we solve the following models, one per each $\overline{Z}^n_{jk} > 0$. Solving these problems generates dual variables λ_{jko} which provide a stronger cut (equation (28)) to add to formulation (REQ). Let $J^= = \{j \in J | \overline{Z}^n_{jk} = 0\}$ and $K^= = \{k \in K | \overline{Z}^n_{jk} = 0\}$. Formulation (R2-SP(0)) presented below is solved to find these optimal solutions.

$$D^=(\overline{Z}^n, o) = \max_{\lambda} : - \sum_{j \in J^=} \sum_{k \in K^=} \nu_{jk} \lambda_{jko}$$

Subject to:

$$\begin{aligned}
g_{ik}\overline{\beta}_o - \overline{\alpha}_{io} - g_{ik}\lambda_{jk} &\leq \bar{c}_{ijk}(o) & \forall i \in I, j \in J^=, k \in K^= \quad (36)\\
\lambda_{jk} &\in R^+ & \forall j \in J^=, k \in K^=. \quad (37)
\end{aligned}$$

The optimal solution to models (R2-SP(0)) can as well be found by inspection. Constraints (36) can be written as $\lambda_{jk} \geq \overline{\beta}_o - \left(\frac{\overline{\alpha}_{io} + \bar{c}_{ijk}(o)}{g_{ik}}\right), \forall i \in I, j \in J^=, k \in$

$K^{=}$. Let $r_{jk} = \max_{i \in I} \overline{\beta}_o - \left(\frac{\overline{\alpha}_{io} + \overline{c}_{ijk}(o)}{g_{ik}}\right)$. Then, the optimal $\overline{\lambda}_{jk}$ for model (R2-SP(0)) can be calculated as

$$\overline{\lambda}_{jk} = \begin{cases} r_{jk} & if \quad r_{jk} > 0 \\ 0 & \text{o/w.} \end{cases} \tag{38}$$

Solving the dual of (SP(o)) using the two-phase procedure enhances the solution time of the L-shaped algorithm. This is mainly due to generating stronger cuts. Additionally, the procedure proposed significantly reduces the size of the LPs solved during Phase 1, which enables us to solve large problem instances without running into memory problems due to using commercial software packages to solve these LPs.

4.1.4. Other algorithmic improvements:

Formulation (REQ) is an integer linear program, as such, it is difficult to solve. During the initial iterations of the L-shaped algorithm, the quality of the solutions obtained from solving (REQ) is poor. This is due to the fact that, in the beginning of the algorithm, the number of cuts (28) added to the (REQ) formulation is very small. The quality of the solutions found improves as the algorithm progresses. For this reason, in the initial iterations of the algorithm we just seek the first feasible solution to (REQ), rather than the optimal solution. Later on, as the algorithm progresses, we seek high quality solutions to (REQ). Doing this reduces the running time of the L-shaped algorithm by reducing the solution time during the initial iterations. Specifically, we seek only a feasible solution to (REQ) for as long as $(\frac{UB-LB}{UB}) \leq 4\%$. Once this condition is met, we seek solutions which are optimal.

5. Case Studies

5.1. Biomass Supply

The case study developed is focused in the state of Tennessee. Each of its ninety four counties is a biomass supplier. Of these, thirty one counties are considered as potential locations for a biorefinery. We opted for a large number

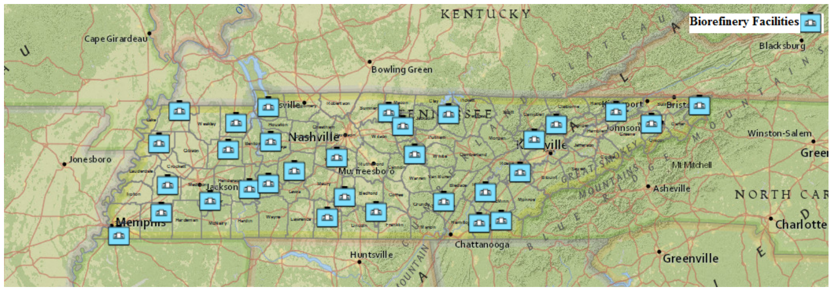

Figure 1: The Biorefineries using the Switchgrass Feedstock Supported by the Counties of the State of Tennessee

of potential biorefinery locations to provide the supply chain with options to reduce costs. The type of biomass considered is switchgrass. Figure 1 illustrates the potential locations of the biorefineries and the counties of the state of Tennessee. The total amount of biomass available for production of bioenergy in the state of Tennessee is 19,482,102.51 dry ton [26].

The amount of switchgrass available at the county level in Tennessee is acquired from the Bioenergy Knowledge Discovery Framework [13]. The amounts reported by the KDF are in dry tons (s_{idry}). Since biomass in the field is not dried, we use equation (39) to convert dried biomass to biomass with a certain moisture content. Moisture content depends on the weather conditions under a particular scenario. The moisture content and the moist biomass supply are computed based on the region the supplier is in and for each year (see Table 4). A random number is generated to determine the moisture content from either the lower (case 1) or upper portion (case 2) of the triangular distribution (i.e., e_i belongs to $at_i \leq e_i \leq t_i$ or $t_i \leq e_i \leq bt_i$). Lastly, equation (39) is used to compute the amount of biomass available at each supplier during 2004 to 2014.

$$s_i = s_{idry}/(1 - e_i) \tag{39}$$

5.2. Biomass Quality and Transportation Costs

The data on the physical and chemical composition of switchgrass came from other studies in the literature. For example, [51] collected samples from bales which were stored up to 75, 150, and 225 days in the state of Tennessee. These samples were analyzed to characterize the moisture and ash content of biomass. The samples were collected from bales with different particle size and wrapping materials which represent different practices that affect biomass quality. Specifically, the data collected was divided into three levels, corresponding to the particle size of full-length of 243.84 cm, 7.62 cm and 1.27-1.91 cm. We use this data to derive the corresponding distributions of moisture and ash content in switchgrass. The goodness of fit test indicated that triangular is a suitable distribution to represent these two random variables. Tables 2 and 3 show the parameters of the triangular distributions at each level for the moisture and ash content, respectively. These distributions characterize the moisture and ash content expected when three different harvesting practices are adopted.

In order to estimate the cost of quality, the values of α'_i, β'_i, α'_i and α_i, β_i, α_i were computed based upon the derivations in Appendix A. The procedure developed to obtain the coefficient p and q (necessary to compute the quality costs due to ash and moisture contents, $c'_i(t_k, o)$ and $c_i(\delta_k, o)$) is described below. Tables 10 and 11 in Appendix A summarize the unit cost of quality used in our numerical analysis.

[18] used experimental data to derive a linear relationship between the percentage of ash content in biomass and the percentage of total liquid yield [mf wt]. We use this linear relationship to compute the percentage of oil yield when ash content falls within 1% and 8%. Next, we calculated the oil yield (in gal/dry ton) assuming an 84 gal/dry ton yield from biomass that contains 1% ash as a baseline. Finally, we used the price of heating oil (at \$1.76/gal on June 2015) as a proxy for pyrolysis oil to calculate the cost of oil per dry ton of biomass feedstock. A penalty cost is charged to biomass feedstock with ash concentrations exceeding the target value of δ_k. This cost equals \$0 at or below the target level (δ_k). The following equation estimates the penalty cost for ash content higher than the technology target level for the thermochemical process (k=1) which

Table 2: The Parameters of Triangular Distribution for the Moisture Content

Division	at	t	bt
Moisture Content (in %) - Distribution 1	26	27	29
Moisture Content (in %)- Distribution 2	17	19	20
Moisture Content (in %)- Distribution 3	16	18	23

Table 3: The Parameters of Triangular Distribution for Ash Content

Division	$c\delta$	δ	$d\delta$
Ash Content (in %)- Distribution 1	1.33	2.89	4.53
Ash Content (in %)- Distribution 2	0.71	2.44	3.79
Ash Content (in %)- Distribution 3	0.82	2.18	3.49

requires ash content to be $\delta_1 \leq 1\%$.

$$m = 5.8561 + 0.6507(\delta - \delta_1)^2 \tag{40}$$

In the case of the biochemical process ($k = 2$), [27] estimate the cost of ash dockage to be \$2.25/dry T per each percentage of ash above the 5% process specification. We consider this to be the unit cost necessary to meet the required ash specifications. Thus, in our model the losses associated to ash (with $\delta_2 = 5\%$) are:

$$m = 2.25(\delta - \delta_2)^2 \tag{41}$$

Noteworthy our estimated distributions are below the 5% specification; therefore, the ash quality cost for this case study is negligible. However, distribution 1 includes the 4% ash content for switchgrass assumed in [27], confirming that this case study is realistic and that the reason why the ash quality cost is negligible is the less stringent process requirements of the biochemical technology.

The expected quality cost associated to moisture is q (see equation (4)). This cost occurs when biomass is mechanically dried to meet process specifications [35]. [35] estimate the fixed cost of drying to be \$2.46/dry ton. They also estimate an additional operational cost of \$7.84/dry ton for an initial moisture content of 40%. This relationship is used to estimate the quality cost for moisture in the range from 10% to 40%.

According to [28], the moisture content does not have a considerable implication to the biochemical conversion process. However, the moisture content is linked to the degradation and consumption of structural carbohydrates during storage. Thus, the technology specification is generally recognized as 20%. Even when the specification of moisture for the biochemical process is not as stringent as in thermochemical conversion process (10%), the moisture content does impact feedstock transportation costs. The following equation estimates the cost of quality when the moisture content of biomass is higher than the technology target level. The technology level is $t_1 = 10\%$ for the thermochemical processes and $t_2 = 20\%$ for the biochemical processed.

$$n = 5.4318 + 0.0066(t - t_k)^2 \ \forall k = 1, 2 \tag{42}$$

To calculate the transportation costs of biomass, we utilize the following equation presented by [1].

$$c_{ij} = 3.85 + 0.0528 d_{ij} \tag{43}$$

where d_{ij} is the distance from supplier i to facility j in kilometers.

5.3. Production Capacities and the Investment Costs of Biorefineries

Five different biorefinery production capacities are considered in this study. Two technologies are considered: thermochemical and biochemical. Regarding the thermochemical conversion process, the biorefineries are designed to yield 37.8, 75.6, 113.4, 189, and 226.8 million liters of ethanol per year. The investment costs for building the biorefineries with these capacities are estimated to be \$90,850,195, \$145,360,312, \$193,813,750, \$271,339,250, and \$310,102,000. The investment costs are computed considering that, due to the economies of scale, doubling production capacities will increase costs by 60%. For these calculations, the reference capacity is 226.8 million liters per year which requires an investment cost of \$310,102,000 [14]. The investment costs for building a biorefinery which uses the biochemical conversion process are estimated using a similar approach. The reference capacity is 226.8 million liters per year which requires an investment cost of $423,000,000 [24]. The biorefineries are designed to produce an ethanol yield of 37.8, 75.6, 113.4, 189, and 226.8 million liters per year.

In our model formulation, the parameter l_{jk} represent the equivalent annual investment costs for biorefinery located in j and using conversion technology k. In this study we assume a project life of 20 years and an interest rate of 15%. The equivalent annual cost (EAC), which is the cost per year of owning and operating an asset over its entire lifespan, is used to provide the annual cost incurred due to those investments. EAC is calculated using the following equation:

$$EAC = \frac{r(NPV)}{1 - \frac{1}{(1+r)^t}}. \tag{44}$$

Where, NPV stands for net present value, r is the interest rate, and t is the expected lifetime of the project.

The conversion rate (g_{ik}) of switchgrass to cellulosic ethanol via a thermochemical process equals 226.8 liters per dry ton [21]. The conversion rate of switchgrass to cellulosic ethanol via a biochemical process equals 378 liters per dry ton [33]. Furthermore, the total amount of bioethanol demand equals 850 Million liters per year for the state of Tennessee based upon a USDA report on February 2015 [47].

5.4. Generation of Scenarios

Eleven scenarios are created by analyzing the historical precipitation data during 2004 to 2014. This data was obtained from six weather stations of the National Centers for Environmental Information located in the South East, North East, Mid South, Mid North, South West, and North West of Tennessee. Details about these scenarios are provided in Table 4. Historical data was used to calculate the average amount of the precipitation in each region. This average was then compared to the actual precipitation on a particular year. Numbers 1 and 2 in this table are used to show whether precipitation was under or over the average (i.e., case 1 or case 2, respectively).

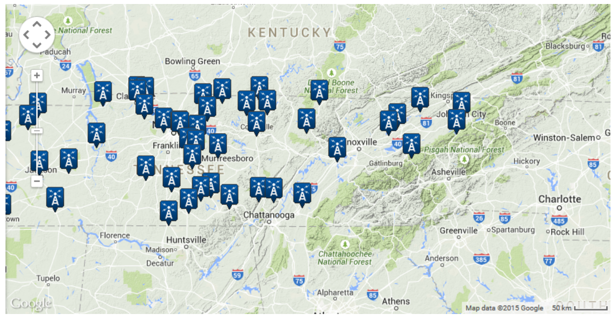

Figure 2: The Stations of State of Tennessee for the National Centers for Environmental Information [37]

Table 4: Scenarios of Regions from 2004 to 2014

Case	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
South East	2	2	1	1	2	2	2	2	2	2	2
North East	1	1	1	1	1	2	1	1	1	1	1
Mid North	2	1	2	2	1	1	1	1	2	1	2
Mid South	2	2	2	1	1	1	2	1	2	1	2
South West	2	1	2	2	2	1	2	2	1	1	1
North West	2	1	2	2	2	2	1	2	2	2	1

We develop six different problems as shown in Table 5. Here, ML-T1, ML-T2, and ML-T3 represent different moisture levels, more specifically, moisture levels following the triangle distributions 1, 2, 3 according to Table 2. Similarly, AL-T1, AL-T2, and AL-T3 represent different ash levels, each following the triangle distributions 1, 2, 3 according to Table 3. The goal of developing these problems is to evaluate how different combinations of ash and moisture content impact the cost of quality in this supply chain.

Table 5: Problem Definitions

Problem	Moisture	Ash
1	ML-T1	AL-T1-Low
2	ML-T1	AL-T1-High
3	ML-T2	AL-T2-Low
4	ML-T2	AL-T2-High
5	ML-T3	AL-T3-Low
6	ML-T3	AL-T3-High

6. Numerical Analysis

This section presents the results of our numerical analysis. The algorithms used in this study are written in AMPL. GUROBI 6.0.0. is used to solve the mathematical models presented above. The experiments are completed using a computer with Intel(R) Core(TM) i7-2600U CPU @ 3.40GHz; and 16.00 GB of RAM.

6.1. Evaluating the Performance of the Stochastic Model

To evaluate the performance of the stochastic programming model proposed we use two performance measures, which are, the value of the stochastic solution (VSS) and the expected value of perfect information (EVPI). Table 6 summarizes the results of this analysis.

VSS represents the cost savings from using a stochastic – instead of the corresponding deterministic – model to design and manage the supply chain. Therefore, VSS is the difference between the objective function value of the stochastic model (REQ) and the expected value (EV) model. We calculate EV by solving (REQ) with a single scenario whose quality cost is equal to the expected value. The value of VSS differs by problem, however, this value is always greater than zero. This fact indicates that the design and management decisions proposed by the stochastic model outperform the decisions proposed by the deterministic model. The savings due to using the stochastic model vary from $19,000 to $59,000 annually.

EVPI measure the value of knowing the future with certainty. We calculate EVPI as the difference between the objective function value of the wait-and-see (WSS) model and the stochastic model (REQ). To find WSS we solve model (REQ) for each scenario assuming that this is the only scenario we will be facing in the future. Next, we use these results to calculate WSS. For example, for Problem 1, $WSS = (1,297,148 * 0.02) + (1,297,540 * 0.03) + ... + (1,297,354 * 0.17) = 1,297,575$. The values of EVPI are positive for all the problems studied. This indicates that, if we were to know the future with certainty, then, the

Table 6: Comparing the Costs (in $1,000) of Stochastic and Deterministic Solutions

Solution Strategies		**Problems**					
		1	**2**	**3**	**4**	**5**	**6**
Wait-and-see							
Sc.	Probability						
1	0.02	1,297,148	1,307,098	1,293,303	1,300,038	1,292,074	1,297,153
2	0.03	1,297,540	1,307,490	1,293,622	1,300,357	1,293,207	1,298,286
3	0.05	1,297,148	1,307,098	1,293,303	1,300,038	1,292,074	1,297,153
4	0.06	1,297,558	1,307,507	1,293,607	1,300,341	1,292,507	1,297,587
5	0.08	1,297,744	1,307,693	1,293,786	1,300,521	1,292,898	1,297,977
6	0.09	1,297,910	1,307,860	1,293,909	1,300,644	1,293,590	1,298,669
7	0.11	1,297,374	1,307,323	1,293,499	1,300,234	1,292,515	1,297,594
8	0.12	1,297,744	1,307,693	1,293,786	1,300,521	1,292,898	1,297,977
9	0.14	1,297,315	1,307,264	1,293,426	1,300,161	1,292,766	1,297,845
10	0.15	1,297,910	1,307,860	1,293,909	1,300,644	1,293,590	1,298,669
11	0.17	1,297,354	1,307,304	1,293,443	1,300,177	1,292,817	1,297,896
WSS		1,297,575	1,307,525	1,293,641	1,300,376	1,292,937	1,298,016
EV		1,297,568	1,307,517	1,293,634	1,300,369	1,292,929	1,298,008
(REQ)		1,297,530	1,307,480	1,293,610	1,300,350	1,292,870	1,297,950
VSS		38	37	24	19	59	58
EVPI		45	45	31	26	67	66

Table 7: List of Algorithms Tested

Algorithm	Description
A1	The L-shaped algorithm presented in Table 1
A2	A1 including improvements presented in Section 4.1.4
A3	A2 including improvements presented in Section 4.1.3
A4	A3 including improvements presented in Section 4.1.1
A5	The multi-cut L-shaped algorithm including improvements presented in Sections 4.1.1, 4.1.3, 4.1.4

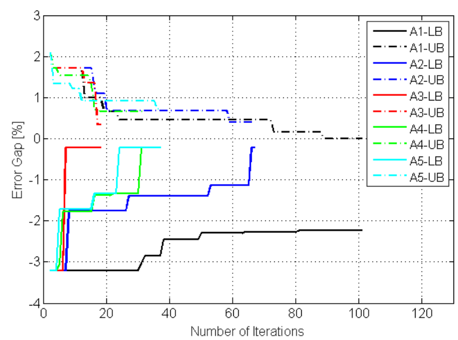

Figure 3: The Performance of the Proposed Five Algorithms over Iterations, including L-shaped, Enhanced L-shaped, and Multi-cut L-shaped Algorithms

performance of this supply chain would be better. For problem 1, for example, the annual cost of knowing the future with certainty is $45K.

6.2. Evaluating Algorithmic Performance

We evaluated the performance of a number of algorithms to solve the problems presented above. Table 7 lists the algorithms developed and Figure 3 illustrates the performance of these algorithms over iterations.

Table 8 summarizes the error gap, running time and the number of iterations for each algorithm. The stopping criteria used for all the algorithms (including Gurobi) was an error gap (i.e., the percentage deviation between the upper and lower bound) less than or equal to 1%. Additionally, we stopped algorithms A1,...,A5 if the total number of iterations reached 100. Based on these results, all the algorithms proposed outperform Gurobi with respect to both, solution quality and running time. The results indicate that, the improvements suggested

Table 8: Summary of Experimental Results

Running time (in CPU sec)

Problem	**Gurobi**	**A1**	**A2**	**A3**	**A4**	**A5**
1	1992.63	100.44	32.92	8.56	16.00	15.92
2	2018.73	96.01	25.80	12.28	6.14	32.92
3	2461.68	105.55	43.44	36.68	29.22	82.79
4	2507.06	111.60	26.90	50.20	18.58	30.23
5	2097.93	105.02	5.58	39.63	18.87	65.03
6	2131.74	109.11	21.34	15.84	14.94	100.94
Avg.	2201.63	104.62	26.00	27.20	**17.29**	54.64

Error gap (in %)

Problem	**Gurobi**	**A1**	**A2**	**A3**	**A4**	**A5**
1	1.00	2.24	0.61	**0.55**	0.87	0.98
2	1.00	2.59	0.86	**0.54**	0.88	0.83
3	1.00	2.66	1.00	0.87	0.78	**0.18**
4	1.00	2.92	0.94	**0.81**	0.84	0.97
5	0.99	2.91	0.89	**0.89**	0.99	0.90
6	0.99	2.88	0.84	0.96	**0.62**	0.81
Avg.	1.00	2.70	0.86	**0.77**	0.83	0.78

Number of iterations

Problem	**Gurobi**	**A1**	**A2**	**A3**	**A4**	**A5**
1	-	100	67	18	32	32
2	-	100	55	28	12	37
3	-	100	75	72	57	77
4	-	100	51	86	38	45
5	-	100	13	73	41	62
6	-	100	42	42	35	74
Avg.	-	100.00	50.50	53.17	**35.83**	54.50

do impact the performance of the L-Shaped algorithm.

The results indicate that A4 was the first to meet both stoping criteria. Its average running time was 17.29 CPU seconds. In terms of solution quality, algorithm A3 outperformed the rest. The quality of the solutions from the multi-cut L-shaped algorithm is also very good.

6.3. Discussing Managerial Insights

Table 9 provides details about the distribution of the total costs for the 6 problems solved. Based on these results, on average, the cost of lowering moisture content to meet process requirements counts for 1.85% of the total costs. The cost occurred due to ash content being higher than process requirement counts for 2.19% of the total costs. On the average, the quality-related costs count for 4.04% of the total supply chain costs. While these costs represent only a small percentage of the overall supply chain costs, in absolute terms, they equal \$52.5 million annually. As such, should not be ignored in the supply chain decision making process.

The quality costs are the highest in problems where moisture and ash levels follow distribution 1 (Problems 1 and 2). These costs are 24.03% and 23.86% higher as compared to the problems where moisture and ash levels follow distributions 2 and 3, respectively. Additionally, ash costs for problems 1, 3 and 5 (which have low ash content as shown in Table 5) are always lower as compared to 2, 4 and 6 (which have high ash content as shown in Table 5). These results indicate that high moisture and ash contents negatively impact the quality of biofuel, and consequently, costs in the supply chain. Indeed, variable costs (i.e., transportation, ash and moisture) increase by 16.5% when comparing problem 2 (highest ash and moisture) with problem 5 (lowest ash and moisture). Moreover, facility locations are also different in problems 1 and 2 versus 3 to 6. In problems 1 and 2, a large capacity plant is located in Giles county and a smaller plant is located in Haywood county. Whereas, in problems 3 to 6, a smaller plant capacity is located in Giles county, and a larger plant is located in Haywood county.

6.4. Impact of Quality Costs on the Supply Chain Network Design and Management

The impact of quality costs on the supply chain network design and management is illustrated in Figures 4 and 5. Figure 4 presents the solution of problem 2 when solving the model which captures quality-related costs. Figure 5 presents the solution of problem 2 when solving a model which does not capture quality-related costs. The model which captures quality costs is mindful of the biomass specifications in certain regions, which impacts the location of plants. The results show that the optimal location of biorefineries differ when biomass quality is included or excluded from the model. Specifically, the model which captures the cost of quality opens a fourth biorefienery in Henry county

Figure 4: Solution using the Highest Parameters in the Distribution for Moisture and Ash in the State of Tennessee when integrating quality-related costs.

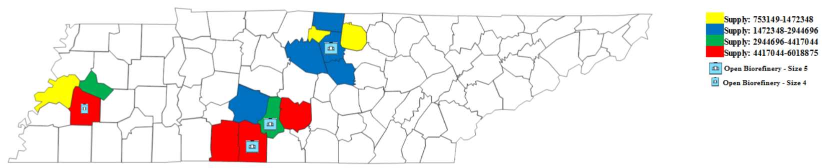

Figure 5: Solution using the Highest Parameters in the Distribution for Moisture and Ash in the State of Tennessee when not integrating quality-related costs.

while the model which does not include the cost of quality opens a fourth biorefinery in Smith county. Interestingly, when considering biomass quality, the model tends to open two biorefineries next to each-other in order to reduce the quality-related costs. This suggests that scalability of plant capacity might be part of the optimal decision. Moreover, biomass transportation paths differ as displayed using the colorbar in Figures 4 and 5. These results also point to the fact that quality-related costs incur when using field run biomass, therefore, they should not be ignored in the supply chain decision making process.

The moisture and ash costs per technology shown in Tables 10 and 11 in the Appendix A were computed using the baseline distributions of Table 2 and 3.

Table 9: Summary of Experimental Results

	Costs (in $ mill)						**Biomass used**
Problem	**Fixed**	**Transport.**	**Moisture**	**Ash**	**Variable**	**Total**	(in mill tons)
1	1,202	41	28	27	96	1,298	3.75
2	1,202	41	28	37	106	1,307	3.75
3	1,202	46	22	24	92	1,294	3.75
4	1,202	46	22	31	99	1,300	3.75
5	1,202	45	22	23	91	1,293	3.75
6	1,202	45	22	29	96	1,298	3.75

7. Conclusions and Future Work

This paper proposes a two-stage stochastic programming model and provides a unique approach to capture the impacts of biomass quality and variability on the design and management of its supply chain. The proposed model minimizes the total of location, transportation, technology selection, and quality costs in the supply chain. To solve the proposed stochastic model, different algorithms were proposed and tested (i.e., L-shaped, L-shaped with trust region cuts and algorithmic improvements, and multi-cut L-shaped algorithm). All of the proposed algorithms outperform Gurobi in terms of solution quality and running time. The results show that A4 is the fastest algorithm with an average running time of 17.29 CPU seconds and 0.83% error gap. Algorithm A3 outperforms the rest in terms of solution quality with an average error gap of 0.77% and an average running time of 27.20 CPU seconds.

The results indicate that the moisture-related cost is 27% higher in Problems 1 and 2 for which the moisture content is highest. The ash-related cost is 31% higher for problems 2, 4 and 6 for which ash content is highest. This highlights the impact of ash and moisture contents on the supply chain costs. Moreover, the results illustrate that high moisture and ash contents negatively impact the quality of biofuel and require the addition of quality control activities; consequently, variable costs (transportation, ash and moisture). For this reason, costs were 16.5% higher for problem 2 (highest ash and moisture) versus problem 5 (lowest ash and moisture).

We also resolved the (REQ) model after dropping the quality-related costs. Hiding the quality-related costs decreased the variable costs, transportation costs included. Moreover, the optimal solution of this model (i.e., location of biorefineries, the selection of suppliers and the amount of biomass transported) differs from the solution when the model considers the quality costs. Since the quality-related costs are incurred when using field run biomass, they should not be ignored in the supply chain decision making process.

The proposed model can be extended to include other types of biomass feedstocks and additional emerging biomass conversion technologies. Future lines of research include the extension of this model to a multi-objective optimization model which minimizes costs and the environmental impacts of biofuel production process. This model can be also extended to further investigate the feasibility of large-scale, region-based supply chains in support of biofuel production. In this case, a biofuel plant would use local suppliers and suppliers located further away to replenish biomass inventories. Expanding the supplier base will, in return, increase biomass availability, increase biomass variety, and reduce biomass supply risks. In order to facilitate the delivery of biomass, a hub-and-spoke biomass SC design model can be investigated.

8. Acknowledgments

This material is based upon work supported by the U.S. Department of Agriculture/National Institute of Food and Agriculture (2015-38422-24064) and U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office (4000142556).

Appendices

Appendix A Calculating Expected Costs for $(\epsilon - t_k)^2$

Case 1: Moisture content of supplier i is lower than t_i, thus, $at_i \leq e_i \leq t_i$. The corresponding density function for e_i is:

$$f_{\epsilon(t)}(e) = \begin{cases} \frac{2(e-at)}{(t-at)^2} & at \leq e \leq t \\ 0 & \text{o/w} \end{cases} \tag{45}$$

$$\phi_1(t_i, \epsilon_i) = \int_{at_i}^{t_i} \frac{2m(\epsilon_i - t_k)^2(\epsilon_i - at_i)}{(t_i - at_i)^2} d\epsilon_i = \frac{2m}{(t_i - at_i)^2} \int_{at_i}^{t_i} (\epsilon_i - t_k)^2 (\epsilon_i - at_i) d\epsilon_i =$$

$$\frac{2m}{t_i^2(1-a)^2} \int_{at_i}^{t_i} (\epsilon_i - t_k)^2 (\epsilon_i - at_i) d\epsilon_i =$$

$$\frac{2m}{t_i^2(1-a)^2}\left(\frac{1}{4}t_i^4-\frac{1}{4}a^4t_i^4+\frac{1}{3}(-2t_k-at_i)(t_i^3-a^3t_i^3)+\frac{1}{2}(t_k^2+2t_kat_i)(t_i^2-a^2t_i^2)-t_k^2at_i(t_i-at_i)\right)=$$

$$2m\left(t_k^2\frac{\frac{1}{2}a^2-a+\frac{1}{2}}{(1-a)^2}+t_k\frac{(-\frac{2}{3}-\frac{a^3}{3}+a)t_i}{(1-a)^2}+\frac{(\frac{1}{4}+\frac{1}{12}a^4-\frac{a}{3})t_i^2}{(1-a)^2}\right)=$$

$$m\left(t_k^2\frac{a^2-2a+1}{(1-a)^2}+t_k\frac{(-4-2a^3+6a)t_i}{3(1-a)^2}+\frac{(6+2a^4-8a)t_i^2}{12(1-a)^2}\right)=$$

$$m\left(t_k^2(\frac{1-a}{1-a})-t_k\frac{(4-2a(1+a))t_i}{3(1-a)}+\frac{(3-a(1+a+a^2))t_i^2}{6(1-a)}\right)=$$

$$m\left(t_k^2-\frac{(4+2a)t_i}{3}t_k+\frac{(3+2a+a^2)t_i^2}{6}\right)$$

In this case:

$$\gamma_i'=1,\qquad \beta_i'=\frac{2(2+a)t_i}{3},\qquad \alpha_i'=\frac{(3+2a+a^2)t_i^2}{6}.$$

Case 2: Moisture content of supplier i is higher than t_i, thus, $t_i<e_i\leq bt_i$. The corresponding density function for e_i is:

$$f_{\epsilon(t)}(e)=\begin{cases}\frac{2(bt-e)}{(bt-t)^2} & t\leq e\leq bt\\ 0 & \text{o/w}\end{cases}\tag{46}$$

$$\phi_1(t_i,\epsilon_i)=\int_{t_i}^{bt_i}\frac{2m(\epsilon_i-t_k)^2(bt_i-\epsilon_i)}{(bt_i-t_i)^2}d\epsilon_i=\frac{2m}{t_i^2(b-1)^2}\int_{t_i}^{bt_i}(\epsilon_i-t_k)^2(bt_i-\epsilon_i)d\epsilon_i=$$

$$=\frac{2m}{t_i^2(b-1)^2}\int_{t_i}^{bt_i}\left(bt_i\epsilon_i^2-\epsilon_i^3-2bt_it_k\epsilon_i+2t_k\epsilon_i^2+bt_it_k^2-t_k^2\epsilon_i\right)d\epsilon_i=$$

$$=\frac{2m}{t_i^2(b-1)^2}\left(\frac{-\epsilon_i^4}{4}+bt_i\frac{\epsilon_i^3}{3}+2t_k\frac{\epsilon_i^3}{3}-bt_it_k\epsilon_i^2-t_k^2\frac{\epsilon_i^2}{2}+bt_it_k^2\epsilon_i\right)_{t_i}^{bt_i}=$$

$$= \frac{2m}{t_i^2(b-1)^2}\left(\frac{1}{4}t_i^4-\frac{1}{4}b^4t_i^4+\frac{1}{3}(2t_k+bt_i)(b^3t_i^3-t_i^3)+\frac{1}{2}(-t_k^2-2t_kbt_i)(b^2t_i^2-t_i^2)+t_k^2bt_i(bt_i-t_i)\right) =$$

$$= \frac{2m}{t_i(b-1)}\left(-\frac{1}{4}t_i^3(b+1)(b^2+1)+\frac{1}{3}(2t_k+bt_i)(b^2t_i^2+bt_i^2+t_i^2)+\frac{1}{2}(-t_k^2-2t_kbt_i)(bt_i+t_i)+t_k^2bt_i\right) =$$

$$= \frac{2m}{t_i(b-1)}\left[\left(t_k^2bt_i - \frac{t_k^2}{2}(bt_i+t_i)\right) + \left(\frac{2t_k}{3}(b^2t_i^2+bt_i^2+t_i^2) - t_kbt_i(bt_i+t_i)\right) + \left(-\frac{1}{4}t_i^3(b+1)(b^2+1) + \frac{bt_i}{3}(b^2t_i^2+bt_i^2+t_i^2)\right)\right] =$$

$$= \frac{2m}{t_i(b-1)}\left[\left(\frac{bt_i-t_i}{2}\right)t_k^2 + \left(\frac{t_i^2}{3}(1-b)(2+b)\right)t_k + \left(\frac{t_i^3(b-1)(b^2+2b+3)}{12}\right)\right]$$

$$= mt_k^2 - m\left(\frac{2t_i}{3}(2+b)\right)t_k + m\left(\frac{t_i^2(b^2+2b+3)}{6}\right)$$

In this case:

$$\gamma_i' = 1, \qquad \beta_i' = \frac{2(2+b)t_i}{3}, \qquad \alpha_i' = \frac{(3+2b+b^2)t_i^2}{6}.$$

Similarly, one could derive the expected quality loss function due to ash content.

735 **Case 1:** Ash content of products from supplier i is lower than δ_i, thus, $a\delta_i \le v_i \le \delta_i$.

$$\gamma_i = 1, \qquad \beta_i = \frac{2(2+a)\delta_i}{3}, \qquad \alpha_i = \frac{(3+2a+a^2)\delta_i^2}{6}.$$

Case 2: Ash content of supplier i is higher than δ_i, thus, $\delta_i < v_i \le b\delta_i$.

$$\gamma_i = 1, \qquad \beta_i = \frac{2(2+b)\delta_i}{3}, \qquad \alpha_i = \frac{(3+2b+b^2)\delta_i^2}{6}.$$

Table 10: The Moisture Cost and Ash Cost for Thermochemical Technology [dollars/ton]

	$c_i(\delta_1, o)$ of case 1	$c_i(\delta_1, o)$ of case 2	$c'_i(t_1, o)$ of case 1	$c'_i(t_1, o)$ of case 2
Distribution 1	7.1719	9.8267	7.2655	7.4932
Distribution 2	6.4493	8.2463	5.8916	6.0071
Distribution 3	6.2702	7.6254	5.8220	6.0992

Table 11: The Moisture Cost and Ash Cost for Biochemical Technology [dollars/ton]

	$c_i(\delta_2, o)$ of case 1	$c_i(\delta_2, o)$ of case 2	$c'_i(t_2, o)$ of case 1	$c'_i(t_2, o)$ of case 2
Distribution 1	0	0	5.7255	5.8212
Distribution 2	0	0	5.4516	5.4351
Distribution 3	0	0	5.4700	5.4392

References

[1] Abbas, D., Handler, R., Dykstra, D., Hartsough, B., & Lautala, P. (2013). Cost analysis of forest biomass supply chain logistics. *Journal of Forestry*, *111*, 271–281.

[2] Ahmed, S., Tawarmalani, M., & Sahinidis, N. (2004). A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming*, *100*, 355–377.

[3] Arora, S. M. (2014). Advanced Biofuels: Creating jobs and lower prices at the pump. URL: http://www.ag.senate.gov/hearings/advanced-biofuels_creating-jobs-and-lower-prices-at-the-pump.

[4] Berndes, G., Hoogwijk, M., & van den Broek, R. (2003). The contribution of biomass in the future global energy supply: a review of 17 studies. *Biomass and Bioenergy*, *25*, 1–28. URL: http://linkinghub.elsevier.com/retrieve/pii/S096195340200185X. doi:10.1016/S0961-9534(02)00185-X.

[5] Birge, J. R., & Louveaux, F. V. (1988). A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, *34*, 384–392.

[6] Birge, J. R., & Louveaux, F. V. (1997). *Introduction to stochastic programming*. Springer, New York.

[7] Caroe, C. C., & Schultz, R. (1999). Dual decomposition in stochastic integer programming. *Operations Research Letters*, *24*, 37–45.

[8] Caroe, C. C., & Tind, J. (1998). L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming*, *83*, 451–464.

[9] Chen, C. W., & Fan, Y. (2012). Bioethanol supply chain system planning under supply and demand uncertainities. *Transportation Research Part E*, *48*, 150–164.

[10] Cundiff, J. S., Dias, N., & Sherali, H. D. (1997). A linear programming approach for designing a herbaceous biomass delivery system. *Bioresource Technology*, *59*, 47–55.

[11] Curry, G., & Deuermeyer, B. (1986). Optimal scheduling of multiple feedstock batch biogas production systems optimal scheduling of multiple feedstock batch biogas production systems. *IIE Transactions*, *18*.

[12] Dauber, J., Brown, C., Fernando, A. L., Finnan, J., Krasuska, E., Ponitka, J., Styles, D., Thrän, D., Van Groenigen, K. J., Weih, M. et al. (2012). Bioenergy from surplus land: environmental and socio-economic implications. *BioRisk*, *7*, 5.

[13] Department of Energy, U. (2011). *U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry*. Technical Report August Oak Ridge National Laboratory Oak Ridge, TN.

[14] Eksioglu, S. D., Acharya, A., Leightley, L. E., & Arora, S. (2009). Analyzing the design and management of biomass-to-biorefinery supply chain. *Computers & Industrial Engineering*, *57*, 1342–1352.

[15] Eksioglu, S. D., & Klein, C. M. (2015). Supply Chain Management of Biomass Feedstock. In S. J. Bhaskar, & T. (Eds.), *Biomass and Biofuels: Advanced Biorefineries for Sustainable Production and Distribution*

(p. 384). CRC Press. URL: http://www.crcpress.com/product/isbn/9781466595316.

[16] EPA (2013). Environmental protection agency proposes carbon pollution standards for new power plants. *https://www.epa.gov/cleanpowerplan*, .

[17] Escudero, L. F., Garin, A., Merino, M., & Perez, G. (2009). A general algorithm for solving two-stage stochastic mixed-integer programs. *Computers & Operations Research*, *36*, 2590–2600.

[18] Fahmi, R., Bridgwater, A., Donnison, I., Yates, N., & Jones, J. (2008). The effect of lignin and inorganic species in biomass on pyrolysis oil yields, quality and stability. *Fuel*, *87*, 12301240.

[19] Gebreslassie, B. H., Yao, Y., & You, F. (2012). Design under uncertainty of hydrocarbon biorefinery supply chains: Multiobjective stochastic programming models, decomposition algorithm, and a comparison between CVaR and downside risk. *AIChE Journal*, *58*, 2155–2179.

[20] Hamming, R. (1950). Error detecting and error correcting codes. *Bell Syst. Tech. J.*, *9*, 147160.

[21] Haque, M., & Epplin, F. (2010). Switchgrass to ethanol: A field to fuel approach. *the Agricultural and Applied Economics Association*, .

[22] Hemmecke, R., & Schultz, R. (2001). Decomposition methods for two-stage stochastic integer programs. chapter Online Optimization of Large Scale Systems. (pp. 601–622). Springer.

[23] Huang, Y., Chen, C. W., & Fan, Y. (2010). Multistage optimization of the supply chains of biofuels. *Transportation Research Part E*, *46*, 820–830.

[24] Humbird, D. R. T. L. K. C. H. D. A. A. S. P. L. J. O. B. W. M. S. D. D. D., D. (2011). *Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol*. Technical Report National Renewable Energy Laboratory Golden, CO. URL: http://www.nrel.gov/docs/fy11osti/47764.pdf.

[25] Jäger-Waldau, A., & Ossenbrink, H. (2004). Progress of electricity from biomass, wind and photovoltaics in the European Union. *Renewable and Sustainable Energy Reviews*, *8*, 157–182. URL: `http://linkinghub.elsevier.com/retrieve/pii/S136403210300114X`. doi:`10.1016/j.rser.2003.10.003`.

[26] KDF (2015). Knowledge discovery framework. us department of energy. *https://bioenergykdf.net*, .

[27] Kenney, K., Cafferty, K. G., Jacobson, J. J., Bonner, I. J., Gresham, G. L., Smith, W. A., Thompson, D. N., Thompson, V. S., Tumuluru, J. S., & Yancey, N. (2013). *Feedstock Supply System Design and Economics for Conversion of Lignocellulosic Biomass to Hydrocarbon Fuels: Conversion Pathway: Biological Conversion of Sugars to Hydrocarbons The 2017 Design Case*. Technical Report Idaho National Laboratory (INL).

[28] Kenney, K. L., Smith, W. A., Gresham, G. L., & Westover, T. L. (2013). Understanding biomass feedstock variability. *Biofuels*, *4*, 111–127. URL: `http://www.future-science.com/doi/abs/10.4155/bfs.12.83`. doi:`10.4155/bfs.12.83`.

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

[30] Kim, J., Realff, M. J., Lee, J. H., Whittaker, C., & Furtner, L. (2011). Design of biomass processing network for biofuel production using an milp model. *Biomass and bioenergy*, *35*, 853–871.

[31] Laporte, G., & Louveaux, F. V. (1993). The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, *13*, 133–142.

[32] Laporte, G., & Louveaux, F. V. (1993). The integer L-shaped method for

stochastic integer programs with complete recourse. *Operations research letters*, *13*, 133–142.

[33] Laser, M., Larson, E., Dale, B., Wang, M., Greene, N., & Lynd, L. (2009). Comparative analysis of efficiency, enviromental impact, and process economics for mature biomass refining scenarios. *Biofuels, Bioproducts, Biorefining*, *3*, 247–270.

[34] Lewandowski, I. (2015). Securing a sustainable biomass supply in a growing bioeconomy. *Global Food Security*, *6*, 34–42.

[35] Mani, S., Sokhansanj, S., Bi, X., & Turhollow, A. (2006). Economics of Producing Fuel Pellets from Biomass. *Applied Engineering in Agriculture*, *22*, 421–426.

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

[37] NCEI (2015). National centers for environmental information. http://www.ncdc.noaa.gov/cdo-web/.

[38] Roy, T. J. V. (1986). A cross decomposition algorithm for capacitated facility location. *Operations Research*, *34*, 145–163.

[39] Ruszczynski, A. (1997). Decomposition methods in stochastic programming. *Math. Program.*, *79(1)*, 333353.

[40] Santoso, T., Ahmed, S., Goetschalckx, M., & Shapiro, A. (2005). A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, *167*, 96115.

[41] Searcy, E., Flynn, P., Ghafoori, E., & Kumar, A. (2007). The relative cost of biomass energy transport. In *Applied Biochemistry and Biotecnology* (pp. 639–652). Humana Press, Inc. (Vol. 136-1 ed.).

[42] Sen, S., & Higle, J. (2005). c3 theorem and the d2 algorithm for large scale stochastic mixed-integer programming: set convexification. *Mathematical Programming*, *104*, 1–20.

[43] Sen, S., & Sherali., H. (2006). Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Mathematical Programming Ser. A.*, *105*, 203–223.

[44] Shapiro, A., & Philpott, A. (2007). A tutorial on stochastic programming. *Manuscript. Available at www2. isye. gatech. edu/˜ ashapiro/publications. html*, .

[45] Sherali, H., & Zhu, X. (2006). On solving discrete two-stage stochastic programs having mixed-integer first- and second-stage variables. *Mathematical Programming Ser. B.*, *108*, 597–611.

[46] U.S. Department of Energy (2013). Biomass Feedstocks. URL: `http://www1.eere.energy.gov/bioenergy/biomass_feedstocks.html`.

[47] USDA (2015). Fuel ethanol production facilities capacity and utilization rates. http://www.ers.usda.gov/data-products/us-bioenergy-statistics.aspx.

[48] Uster, H., & Agrahari, H. (2011). A benders decomposition approach for a distribution network design problem with consolidation and capacity considerations. *Operations Research Letters*, *39*, 138–143.

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

[50] Yamamoto, H., Fujino, J., & Yamaji, K. (2001). Evaluation of bioenergy potential with a multi-regional global-land-use-and-energy model. *Biomass and Bioenergy*, *21*, 185–203. URL: `http://linkinghub.elsevier.com/retrieve/pii/S0961953401000253`. doi:`10.1016/S0961-9534(01)00025-3`.

[51] Yu, T., Larson, J., English, B., Boyer, C., Tyler, D., & Castillo-villar, K. (2015). Influence of particle size and packaging on storage dry matter losses for switchgrass. *Biomass and Bioenergy*, *73*, 135–144.

[52] Zamboni, A., Shah, N., & Bezzo, F. (2009). Spatially explicit static model for the strategic design of future bioethanol production systems. 1. Cost Minimization. *Energy* & *Fuels*, *23*, 5121–5133.