Chapter 3

Modeling and Optimization of Quality Variability for Decision Support Systems in Biofuel Production

Mario Aboytes-Ojeda and Krystel K. Castillo-Villar

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.73111

Abstract

Biofuels are a promising alternative to fossil fuel depletion, due to their sustainable production from living or recently living organic matter (i.e., biomass). Biofuel production offers benefits that are not present in non-sustainable resources, like the reduction of air pollution. According to government agencies, biofuel production is expected to increase in the U.S. within the next few years because of government initiatives. In order to become a feasible alternative to satisfy market demand, biofuels require strategic improvements in areas such as supply chain management to deal with the variability within the biomass. Advanced analysis tools might be utilized to integrate biomass physical and chemical properties into the decision processes. This chapter introduces a principal component analysis (PCA) to determine significant factors that affect the operations within the supply chain and, later on, incorporates those factors in an optimization model for the decision analysis. The results show that incorporating quality-related properties has a significant impact in the solution of the optimization program.

Keywords: biofuels, biomass, optimization, principal component analysis, stochastic programming, two-stage problems

1. Introduction

Global population is still increasing, and therefore, more resources are required to fulfill people needs. The demand of energy is growing due to the population increase but also because the presence of new activities such as social networking. Nowadays, most of the vehicles utilize fossil fuels. The fossil fuels are non-renewable resources and they contribute significantly to the global pollution.

Biofuels are a propitious alternative to the fossil fuels reliance, due to their sustainable production based on using biomass as a raw material. Biomass is an organic matter derived from living or recently living beings. The production of biofuels also has some advantages from those produced from non-sustainable resources, for instance, the reduction of greenhouse emissions (GHG) such as the CO_2. Another benefit of using biofuels is the development of an agricultural industry and the creation of rural jobs to produce and deliver biomass to biorefineries. Bioethanol is one type of biofuel that can be produced from organic resources such as corn stover, miscanthus, and switchgrass, among others. Bioethanol has several applications in a wide variety of industries. According to government dependencies such as the Energy Information Administration (EIA), an increment of biofuels is expected in the U.S. within the coming years as product of the Renewable Fuel Standard [1] that requires more production of fuel utilizing renewable resources (e.g., 16 billion gallons for 2022).

Biofuels are classified according to the raw material utilized to produce them: (1) the first-generation is related to biofuels produced from edible biomass that can be generally used for human consumption (e.g., corn, sugarcane, sugar beet, among others), (2) the second-generation are biofuels generated from a wide range of feedstock, including lignocellulosic biomass (LCB), such as perennial grasses, soft and hard wood, up to municipal solid waste, and (3) the third-generation commonly refers to biofuels produced from algal biomass [2]. Biofuels produced from LCB are a feasible option in the U.S. for the coming years as they utilize non-food feedstock and can be grown in marginal lands or are byproducts produced from the wood industry.

Biofuel production requires improvements in strategic areas such as conversion technologies, genetic manipulation of feedstock breeds, and supply chain management of biomass from harvesting areas to conversion facilities, among others in order to become a plausible alternative to fossil fuel production. Supply chain (SC) improvement represents an important area of opportunity in biofuel production due to the fact that environmental, geographical and economic factors are related to operations like harvesting, handling, storing as well as transportation, which have shown significant impact in biofuel yields/cost. Studies have demonstrated that some factors such as storage time affects the physical and chemical properties of the biomass [3], therefore, biomass properties have an important role in the design of the operations required for the production and distribution of biofuel [4].

There are many properties that affect the conversion efficiency depending on the type of conversion technology that is being utilized for the production of biofuels. Moisture is a property of biomass that affects both the thermochemical and biochemical conversion technologies and a drying process that diminishes the moisture content in the biomass up to the required level of humidity is needed. The drying process to meet the specification for the selected conversion technology incurs in a cost that could be reduced/controlled with the implementation of logistics processes and infrastructure design that consider the level of humidity in the feedstock. The utilization of the biomass without meeting the expected specifications could lead the production of biofuels to an inefficient conversion process. Another example of the importance of the biomass properties are the carbohydrates. If the level of carbohydrates does not meet the specification, then, the amount of biofuel derived from that specific batch of biomass will be less than expected. With a shortage of biofuel, acquiring the slack from a third-party supplier to cover the demand could lead the producer to an increment in the overall cost.

Biomass has many physical and chemical properties that need to be considered in order to optimize an objective such as minimization of the total cost or maximization of the profit. This chapter introduces a principal component analysis (PCA) to identify significant factors that affect the design and implementation of logistic processes and infrastructure due to the physical and chemical properties of the biomass. Moreover, the chapter presents a two-stage optimization model that take into consideration quality-related costs in order to set up the biorefinery locations and the flows of biomass from the supplier to the producer. The optimization model can be incorporated into a decision supported system (DSS) to solve several instances of interest and aid the decision-making process.

2. A PCA in a switchgrass composition

This analysis focuses on LCB [specifically, on switchgrass (*Panicum virgatum*)], which is a feedstock derived from organic matter that is mainly composed of cellulose and hemicellulose. Examples of LCB are corn stover, wheat straw and switchgrass and there are many activities involved in the production and distribution of this kind of biomass such as harvesting, extracting, packaging, transporting, handling, among others. There are factors in the aforementioned activities that affect the physical and chemical properties of the biomass, and thus, their quality-related costs. The consideration of those factors has an effect in the design and implementation of the supply chain (SC).

An example of the relationship between the biomass properties and the SC design is the cellulose and hemicellulose (carbohydrates) contents in the LCB. The cellulose and hemicellulose content in the LCB is directly related to the biofuel produced in the conversion process since the carbohydrates are the main component to produce the energy. The more carbohydrates contained in the biomass, the more liters of biofuel obtained from that particular batch of feedstock. Hence, activities that affect the carbohydrate contents need to be improved to minimize the impact on the conversion process.

Densification is one of the processes in the SC of biofuel production that affects some of the properties within the biomass. The densification of biomass consists in conglomerating the organic matter in the form of compact structures such as briquettes and pellets, to improve their handling, storage and transportation but also to reduce the level of dry matter loss (DML) in the feedstock. The DML is directly related with the loss of carbohydrates in the biomass. The less organic matter in the batch, the less amount of cellulose and hemicellulose to produce biofuels. Densification also affects other properties in the biomass such as the moisture content, unit density, durability index, as well as other properties specified for the conversion process according to the implemented technology. Controlling the physical and chemical properties under the specification requirements is vital for an efficient conversion. Delivering biomass that does not meet the specifications could lead to extra costs as a consequence of re-processing the biomass up to meeting the specifications.

The reduction of the variance within the physical and chemical properties of the biomass helps to avoid extra operational cost due to re-processing of feedstock that is out of the specifications. Identifying the factors, which are involved in production and distribution activities that affect the physical/chemical properties, is necessary to design an efficient SC. Researchers in the field have studied baling effects in feedstock properties [5–7]. In previous works, Aboytes-Ojeda et al. [3] proposes a PCA to detect those factors that have a significant impact on properties of interest and that should be approached by implementing novel operations and strategies in order to fulfill the conversion specifications.

The multivariate methodology proposed by Aboytes-Ojeda et al. [3] intends to: (1) introduce the covariance information analysis to draw systematic insights about the factors under study, and (2) present a novel methodology to identify the contribution of every factor in the analysis with respect to the total system variability. The variables introduced in the analysis were cellulose, hemicellulose, lignin, ash and extractives content; whereas the factors were the particle size in the bale, the wrap material, the days in storage and the weight of the bale.

2.1. Experimental methodology

In the year 2012, an experiment to find insights related to the physical and chemical properties of switchgrass was designed and implemented at the Biomass Innovation Park in Vonore, Tennessee. The type of biomass was from Alamo switchgrass and the samples were harvested and then baled in squared shape (1.2 m $\times$ 0.9 m $\times$ 2.4 m) with a baler machine New Holland BB9080 (New Holland Agriculture, New Holland, PA, USA). New Holland BB9080 is a square baler without a cutter that was used to process the batch of lignocellulosic biomass in the second week of January of that year. After processing the biomass to form square bales, the biomass was transported to other covered location before the beginning of the pre-processing.

A Vermeer TG5000 tub grinder (Vermeer Corporation, Pella, IA, USA) was utilized to unpack and grind the switchgrass in January 2012. Once the biomass was ground, the next step was to sample the moisture content and the chemical composition; the measures for the chemical composition were obtained with a near-infrared (NIR) technology. A machine BT3 industrial baler (TLA Bale Tech LLC, South Orange, NJ, USA) was employed to bale the switchgrass one more time after measuring the chemical properties. Round bales (1.2 m of diameter $\times$ 1.5 m of width) were made with the BT3 and then they were moved to storage before the next phase of the experiment.

Four controllable factors were introduced in the analysis that utilizes a split-split plot design. The factors in the analysis were: (1) the number of days in storage, (2) the particle size of the feedstock, (3) the wrap type of the bale, and (4) the weight of the bale. The number of days has three groups or levels, same as particle size. The wrap type and the weight of the bale have two groups. The database utilized for this study presented the necessary conditions of normality, homogeneity and heteroscedasticity as discussed in Kline et al. [8].

Factors	No. groups/levels	Variables
Particle size	3	Cellulose, hemicellulose, lignin, ash and extractives
Wrap type	2	Cellulose, hemicellulose, lignin, ash and extractives
Storage days	3	Cellulose, hemicellulose, lignin, ash and extractives
Bale weight	2	Cellulose, hemicellulose, lignin, ash and extractives

Table 1. Factors and variables for PCA.

Table 1 identifies the factors and variables included in the analysis. The storage days were classified in three groups or levels: 75, 150 and 225 days of storage. The particle size was defined in three groups or levels: PS1 (243.84 cm), PS2 (7.62 cm) and PS3 (1.27–1.91 cm). The wrap type was categorized in two levels: (i) net mesh and net (excluding the two ending parts of the bale), and (ii) the high tensile strength film wrapping for the complete bale (net and film). The bale weight has two levels for this study; the lower level was for bale with a weight between 957.65 and 1715.20 lb., whereas the high level was for bale with a weight between 1715.21 and 2455.10 lb. The weight in the bale has repercussions in logistic operations such as handling, storage, and transportation.

Five variables were included in the analysis. The cellulose is a glucose polymer linked by glycosidic bonds and the hemicellulose is a branched polymer of carbon sugars. The lignin refers to a structural component of plants, consisting of an aromatic system made of phenyl proposal units. The ash is considered as the inorganic leftover after the combustion process at 550–600°C. The extractives are non-structural components that can include free sugars, proteins, chlorophyll, and waxes. The PCA methodology proposed uses the variability within the variables (i.e., variance and covariance) to create artificial variables and then it groups the data according to their corresponding factor group/level. Finally, a statistical comparison of means is utilized to conclude if there is a significant difference between the means of every factor group/level. **Figure 1** shows the methodology to perform the PCA which consists of five basic steps.

2.2. Principal component analysis (PCA)

In order to implement the PCA [9, 10], it is necessary to test the required data conditions to perform the analysis, then, the covariance/correlation matrix needs to be calculated. A Bartlett's test of sphericity is utilized next to determine if the correlation information for the analysis is significant. With the covariance/correlation matrix, eigenvalues and eigenvectors are computed. The eigenvalue is utilized to determine the portion of variance attributed to the corresponding eigenvector. The components of the eigenvectors known as loadings are used to transform the original data into the components scores. The variance matrix is shown in **Table 2**; there is no need to compute the correlation matrix since all the measures for every variable are in the same scale.

The portion of variance in PCA is calculated according to the eigenvalues obtained in the analysis. The idea behind the PCA is to detect those components with the higher eigenvalues;

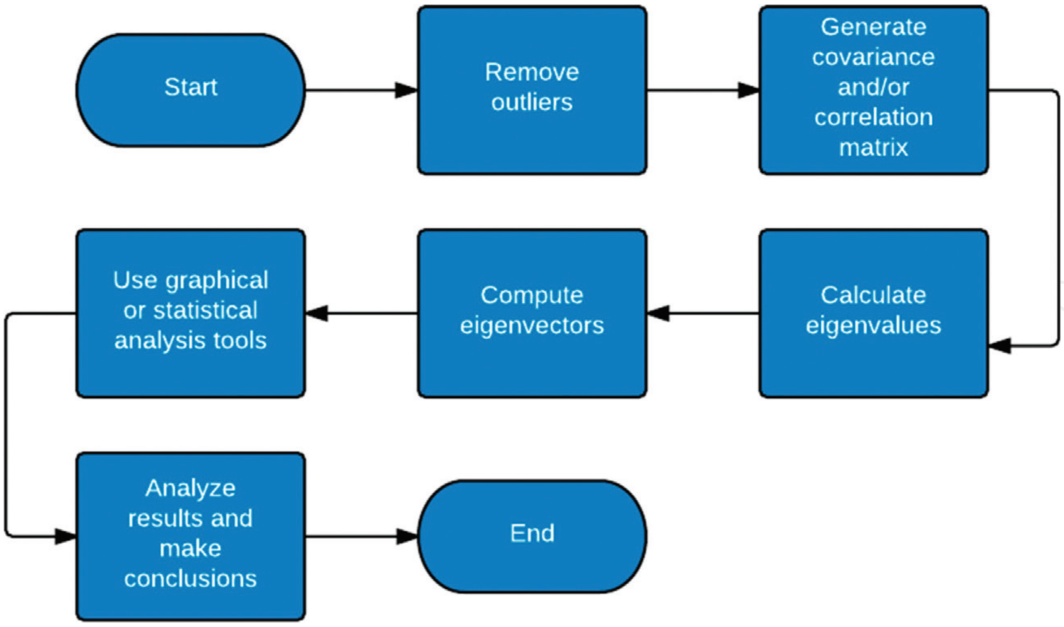


Figure 1. Flow chart for PCA.

Component	Cellulose	Hemicellulose	Lignin	Ash	Extractives
Cellulose	1.344	−0.149	−0.130	−0.325	−0.686
Hemicellulose	−0.149	1.366	−0.467	−0.017	−0.377
Lignin	−0.130	−0.467	0.541	0.198	0.064
Ash	−0.325	−0.017	0.198	0.334	0.226
Extractives	−0.686	−0.377	0.065	0.226	1.100

Table 2. Chemical components covariance.

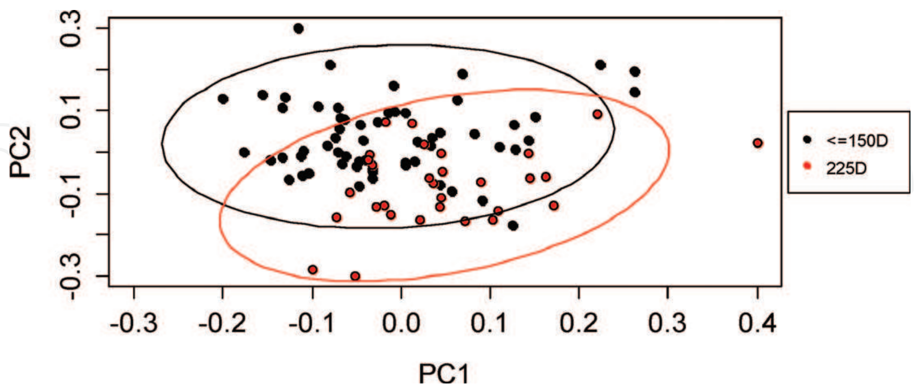


Figure 2. Scree plot for PCA.

therefore, it is possible to identify the principal components due to their share in the total variance. **Figure 2** shows that the first two components contain almost 80% of the total variance; **Table 3** presents the eigenvalues for all the components as well as their share in the total variance. As a rule of thumb, those components above the value of one in **Figure 2** must be considered as the principal components of the analysis.

	Principal component				
	1	2	3	4	5
Eigenvalue	1.448	1.266	0.777	0.527	0.318
Variance (%)	45	34	13	6	2
Cumulative (%)	45	79	92	98	100

Table 3. Variance analysis in PCs.

Figure 2 and **Table 3**, PC1 and PC2 are the main components since their eigenvalues are above one and the amount of variance represents up to approximately 80% of the total variance. The eigenvalues are also attached to the eigenvectors which are the directions where the largest variance is presented. The loadings are the values that indicate the correlation between the original value and the score value. A high value means that the original variable and the component are close. **Table 4** represents the loadings values.

With the loadings values, it is possible to transform the original data into scores. Scores are the representation of the original data points under the principal components basis. These scores can be plotted in a graph called bi-plot and then explored (exploratory data analysis) in order to find some insights related to the segregation within the groups of data. In the results section, several bi-plots are presented to show some of the insights found in the analysis.

Sometimes it is not possible to detect any pattern in the exploratory analysis (i.e., segregation in the data cannot be visually identified). For those occasions, a statistical analysis is needed to determine if there is a significant effect in the principal components due to the factors that were previously introduced. The statistical analysis in this work was performed with a t-test; the means for every group/level within a factor are compared to see if there is any significant difference between them, if so, it can be concluded that there is some evidence to claim that there is a significant effect in the data due to the factors. The statistical test has the following assumptions: unknown but equal population variances, known sample means and not equal sample variances. The following equations are defined for the t-distribution and the corresponding estimator is introduced with the expressions:

$$t = \frac{\overline{x}_1 - \overline{x}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{1}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{2}$$

2.3. Exploratory data analysis and statistical test results

The following set of bi-plot graphs are introduced to show the relevant information about the groups/levels within each one of the factors under study. **Figure 3** presents the data classification according to the wrap type that was utilized to wrap the switchgrass. As it can be

	Principal component variables				
	1	2	3	4	5
Cellulose	−0.64	0.48	0.30	−0.52	−0.09
Hemicellulose	−0.33	−0.83	−0.05	−0.33	−0.32
Lignin	0.21	0.29	−0.66	−0.26	−0.62
Ash	0.22	−0.07	−0.32	−0.64	0.66
Extractives	0.63	−0.02	0.61	−0.39	−0.29

Table 4. Loadings in PCA.

observed, there is no distinguishable segregation within the groups to claim any possible effect due to this factor.

Like the wrap type, data information was also classified according to its particle size and was shown in a bi-plot graph presented in **Figure 4**. The visual representation of the data does not exhibit any clustering in the plotting area, and therefore, no significant findings can be concluded from the bi-plot. Same analysis occurs with the factor corresponding to the classification according to the weight of the bale that can be observed in **Figure 5**, no segregation is noticeable.

In the bi-plot that corresponds to the classification of the date with respect to the days of storage it is possible to identify a segregation in the data. **Figure 6** exhibits this difference between the bales with more than 150 days of storage and those with less number of days. **Figure 7** has another perspective to visually identify this division.

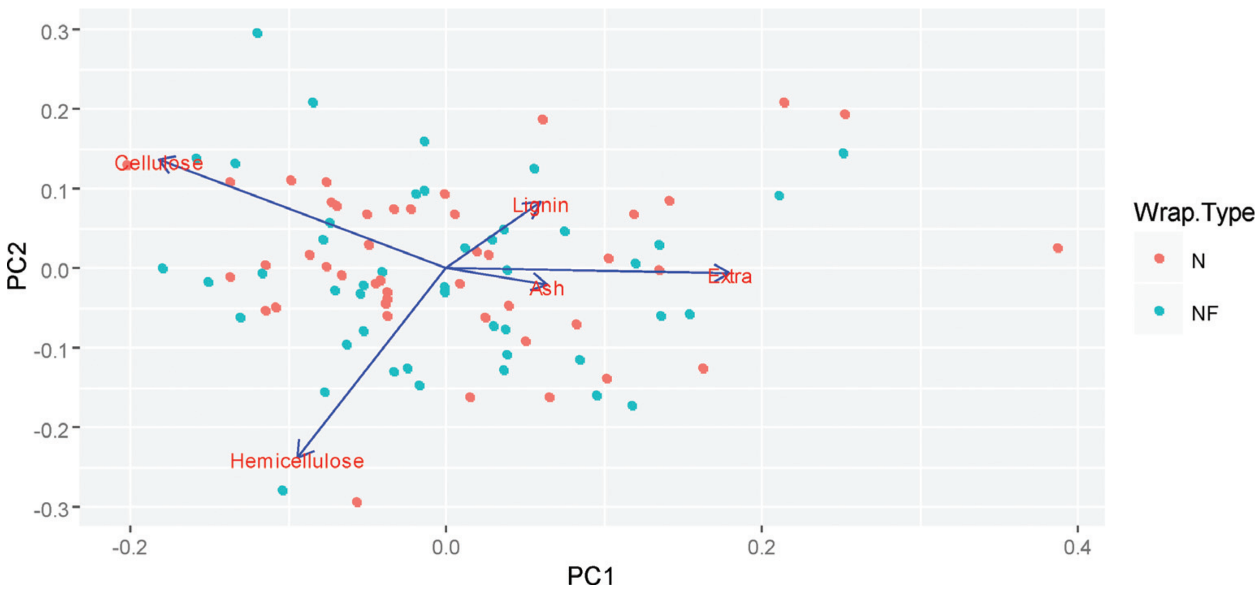


Figure 3. Bi-plot chart for analysis of wrap type.

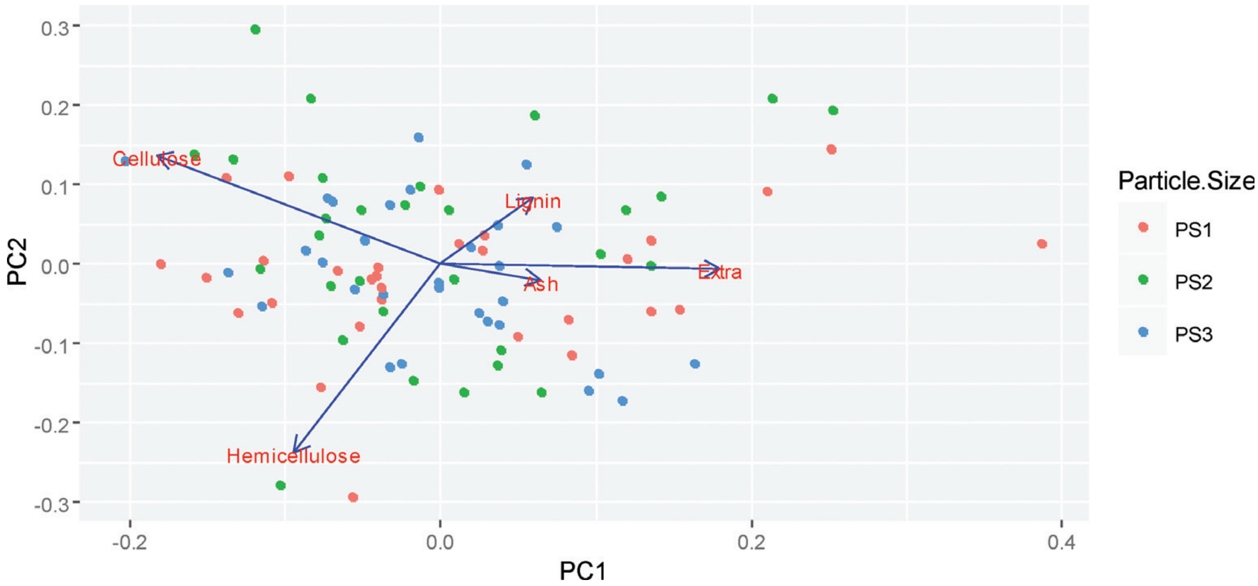


Figure 4. Bi-plot chart for analysis of particle size.

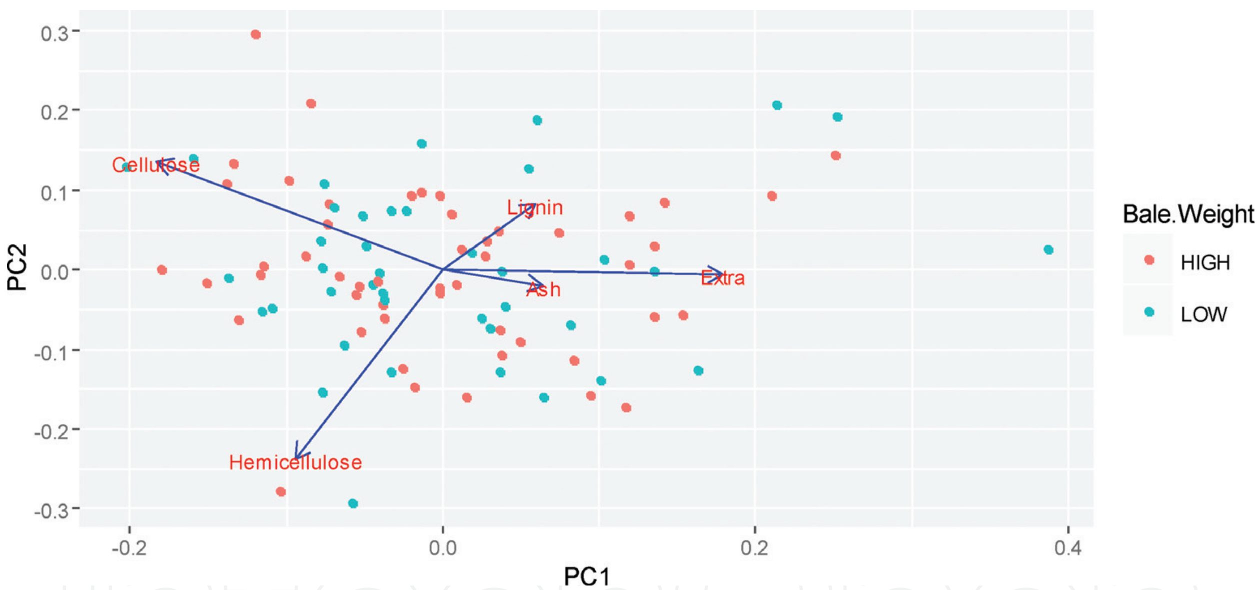


Figure 5. Bi-plot chart for analysis of bale weight.

A t-test was applied for every group/level within every factor presented. The results are shown in **Figure 8**. The storage days is the most important factor since it is the only factor that shows a significant effect due to the statistical difference between the means in the groups. Based on the results, the storage days have repercussion in almost 80% of the variation within the data. Hence, it is relevant in the design of operations that are time dependent.

PCA is a statistical tool that allows the analyst to introduce variance and covariance in the study. Adding the covariance or correlation between the variables could lead the analysis to find some insights that would not be visible with a univariate data analysis tool. Also, PCA allows to sort and classify in a more natural way the significance of every factor analyzed in

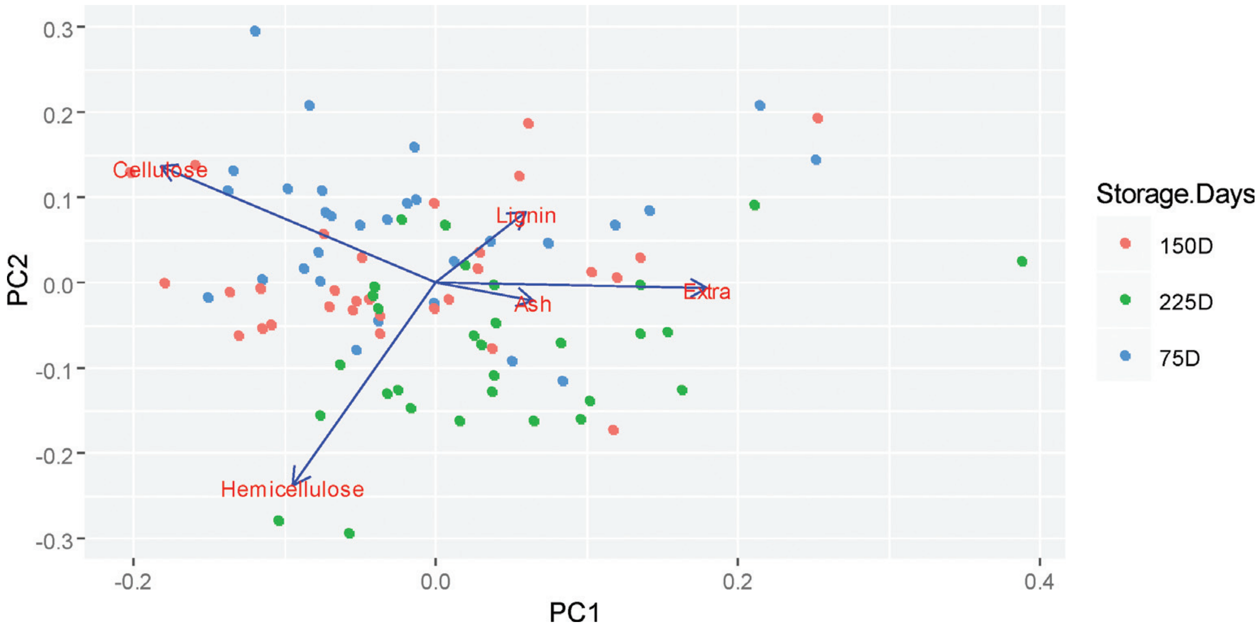


Figure 6. Bi-plot chart for analysis of storage days.

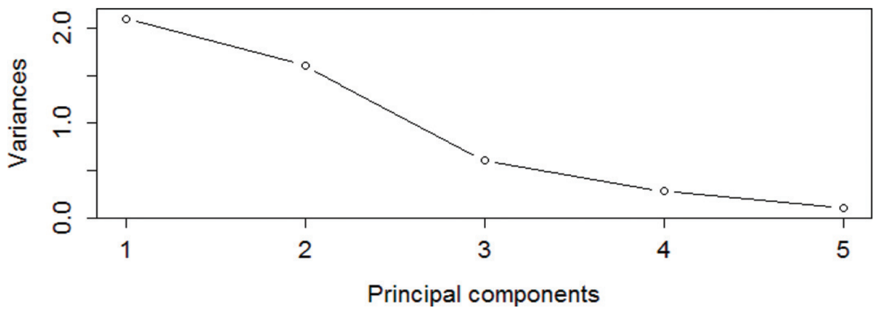
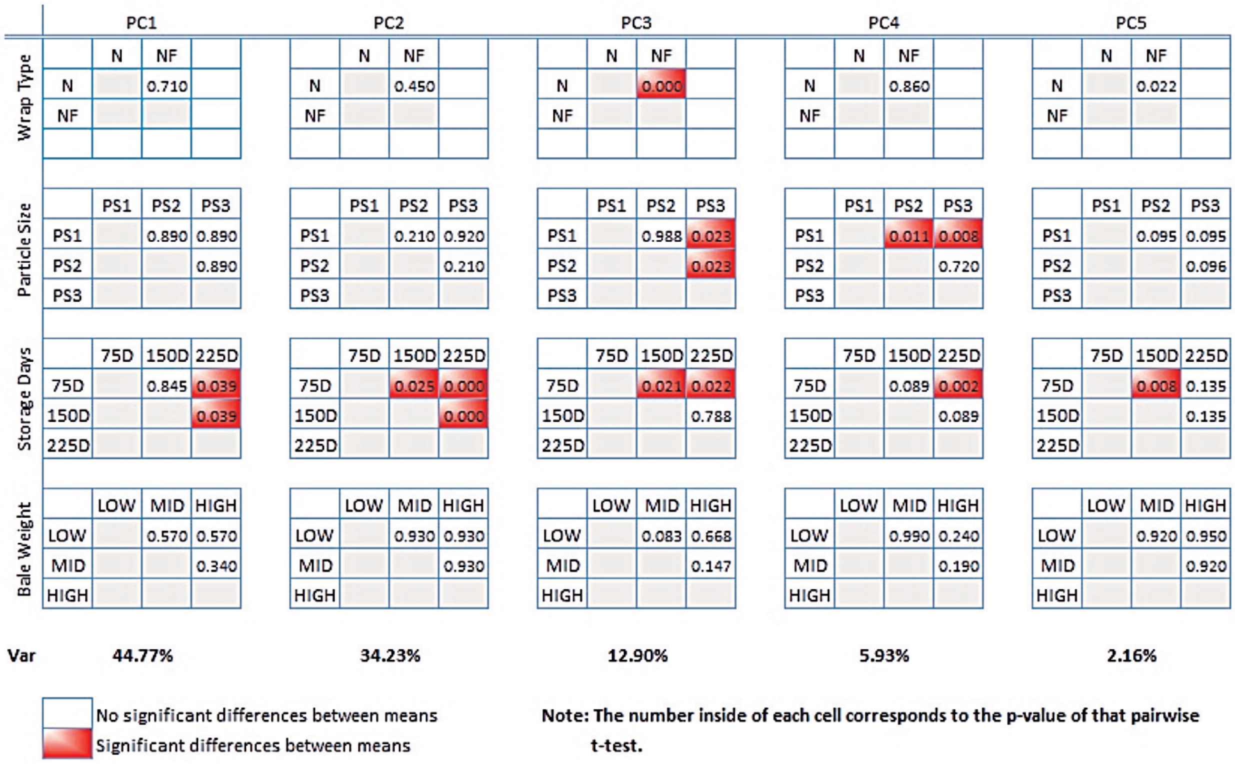


Figure 7. Score plot analysis for PCA.

the database by identifying those factors that have an impact on the main principal components. PCA identifies those factors with a significant effect over the principal components. This information can be utilized to incorporate the relevant factors in stochastic programming models that use the insights found by the multivariate analysis in the optimization problem, and later on, in the decision process. A further discussion on this matter follows.

3. Variable quality-related cost in SC design

The biofuel is an alternative energy resource for the fossil fuels. It is expected that biofuel production increases in the upcoming years due to the increase in the demand. There are some alternatives to approach a mature production of biofuel to support the demand satisfaction. One approach consists in developing new technologies with better conversion processes in order to get an efficient exploitation from the pre-processed biomass for a certain good or



Wrap Type	PC1			PC2			PC3			PC4			PC5		
	N	NF		N	NF		N	NF		N	NF		N	NF	
N		0.710			0.450			0.000			0.860			0.022	
NF															

Particle Size	PC1			PC2			PC3			PC4			PC5		
	PS1	PS2	PS3	PS1	PS2	PS3	PS1	PS2	PS3	PS1	PS2	PS3	PS1	PS2	PS3
PS1		0.890	0.890		0.210	0.920		0.988	0.023		0.011	0.008		0.095	0.095
PS2			0.890			0.210			0.023			0.720			0.096
PS3															

Storage Days	PC1			PC2			PC3			PC4			PC5		
	75D	150D	225D	75D	150D	225D	75D	150D	225D	75D	150D	225D	75D	150D	225D
75D		0.845	0.039		0.025	0.000		0.021	0.022		0.089	0.002		0.008	0.135
150D			0.039			0.000			0.788			0.089			0.135
225D															

Bale Weight	PC1			PC2			PC3			PC4			PC5		
	LOW	MID	HIGH	LOW	MID	HIGH	LOW	MID	HIGH	LOW	MID	HIGH	LOW	MID	HIGH
LOW		0.570	0.570		0.930	0.930		0.083	0.668		0.990	0.240		0.920	0.950
MID			0.340			0.930			0.147			0.190			0.920
HIGH															

Var	44.77%	34.23%	12.90%	5.93%	2.16%

No significant differences between means
Significant differences between means

Note: The number inside of each cell corresponds to the p-value of that pairwise t-test.

Figure 8. T-test results for all the factors.

product. Another approach considers genetic manipulation with different types of biomass to create hybrids with specific characteristics that suit better for a particular final product. Lastly, another alternative is the improvement of the logistics operations related to the production and distribution of the biofuel.

SCs are being used widely in industry as a system of production and distribution due to the need of integrating processes from suppliers until reaching the end users of any good or service. SCs are presented in diverse scales, from local scale, regional scale up to global scale. Nowadays, large-scale are vastly utilized since they reduce operational costs due to the economies of scale. Large-scale SCs are more challenging since they are more difficult to analyze and solve, so they require the development of new algorithms in order to deal with the complexity of the problem. The introduction of models and algorithms to solve large-scale problems is fundamental in industry since real applications require big data, as well as a large set of constraints related to the problem. Bioenergy industry is a field where large-scale SCs are being implemented.

The production and distribution of biofuels are a big challenge since their commercialization requires a competitive price in the market, and therefore, an efficient SC must be implemented to minimize the operational costs. The production of biofuels deals with the inherent variability in the physical and chemical properties within the biomass because of their impact in some key processes such as the conversion process. The second generation of biofuels (e.g., corn stover, miscanthus, and switchgrass) presents, in general, more variation in properties such as moisture and ash content than the first generation, and then, quality-related costs must be contemplated in the design and implementation of the SC. However, one of the advantages of using second-generation biomass for the production of biofuels is that local farmers are familiar with the techniques to cultivate and harvest the biomass. For example, harvesting techniques for forage, utilized to feed the livestock, and for the power grass have similar characteristics.

One common objective of the optimization in SC of biofuels is the minimization of costs under the assumption that all types of biomass have similar properties. This assumption can derive in considering only purchasing, logistics and processing costs which is not adequate since every type of biomass possesses different physical and chemical properties that affect the way the SC works. Not including the aforementioned properties could lead companies to have a negative impact on their expected profit since these properties usually experience high levels of variability. Pilot scale biorefineries have experienced a significant difference between the expected and the actual input of biomass [11]. The randomness in the biomass is one of the challenges that biofuel producers must tackle in order to reach profitability and sustainability.

Biomass conversion technologies have their origins in laboratories where specifications of biomass are controlled. The technologies are designed to work under some specifications within the biomass such as the moisture, ash, and carbohydrates contents. When technologies are implemented in large- scale scenarios, it is very likely to receive biomass with specifications that do not meet the requirements, as a result, additional re-work must be done to utilize it. The quality of biofuel is associated with reaching the target levels of physical and chemical properties, for some specific technology, with a low variability.

A poor quality of the biomass results in higher total costs for companies. A quality-related cost can be defined as any cost derived from not meeting the required specifications for a specific conversion technology. The impact of these quality related costs is usually found after the biorefineries have begun operations. The optimization of biofuels [12, 13] considering randomness in the properties of biomass can be approached with stochastic programming [14, 15]. In literature, most of the supply chain models designed for biofuel production are deterministic; then, the stochastic programming is a novel approach to solve instances with variability in the feedstock.

Novel optimization models take into account inherent properties of biomass to lead better decisions that minimize the total cost. It is important to identify the factors that have an influence over the biomass and include those considerations in the optimization model. Castillo-Villar et al. [16] have proposed a stochastic programming model that includes the quality variability in order to make decisions about important aspects of the SC. The work of Castillo-Villar et al. [16] is revisited in this chapter since the authors present a seminal model that integrated biomass variability in the modeling of biofuel supply chains.

3.1. Quality integration in decision models

Moisture and ash content are important properties of the biomass. Castillo-Villar et al. [16] define a random variable, $\varepsilon(t)$ to represent the moisture content corresponding to the mean value t. A triangular distribution $f_{\varepsilon(t)}$ has been defined for t in the range $[at, bt]$ with a probability density according to the following criteria:

if $at \leq e \leq t$ *then*

$$f_{\in(t)}(e) = \frac{2(e - at)}{(bt - at)(t - at)} \tag{3}$$

if $t < e \leq bt$ *then*

$$f_{\in(t)}(e) = \frac{2(bt - e)}{(bt - at)(bt - t)} \tag{4}$$

Otherwise, 0.

The triangular distribution was proposed due to experimental results in the work of Boyer et al. [17]; in the experiment, a breed of Alamo switchgrass was utilized to test the effect of factors such as storage days, particle size, wrap type and weight of the bale.

Ash content was also represented with a random variable $\vartheta(\delta)$ and the corresponding function is a triangular distribution of the mean value δ. The triangular distribution is defined for the range $[c\delta, d\delta]$. Depending on the technology utilized for the conversion of biomass into biofuel, different requirements are necessary to accomplish the conversion. For example, the conversion of biomass using a thermochemical technology demands at most 10% of moisture content for an efficient process. The moisture target for technology k is defined as t_k. Violating the target for the selected technology will lead to other necessary costs ($\$q$) to compensate not meeting the specifications. The cost of mechanically drying the biomass will be applied to the final good since the content of moisture needs to be reduced up to the target level for conversion purposes. The thermochemical technology also requires at most 10% of ash content for an efficient conversion. The target is defined as δ_k. Again, not meeting the target specification will lead to reprocessing of the biomass with an additional cost.

3.2. Two-stage stochastic model

Stochastic programming introduces randomness into the models where the stochastic variables play a fundamental role in the decision processes. There are several types of stochastic models but probably the most utilized are the two-stage models. In the two-stage models, two types of variables arise, the here-and-now variables and the wait-and-see variables. The here-and-now variables are those that need to be solved in the first-stage because they represent the beginning of the decision and the rest of the decision variables will rely on this step. The wait-and-see variables are presented in the second-stage of the process and depend on the realization of each presented scenario as well as the first-decision stage. The randomness is presented by defining random parameters that will converge to a certain value depending on the scenario, an expected value function for these scenarios in the second-stage plus the value function of the first-stage form the objective function of the stochastic program. A stochastic programming assumes known distributions in order to set the values for the stochastic parameters in every scenario so the program can be maximized or minimized depending on the objective function.

The two-stage stochastic models for location and transportation define the locations of facilities in the first-stage and the transportation of goods in the second-stage. Castillo-Villar et al. [16] defines a stochastic location-transportation model that introduces the location of the biorefinery as the first-stage variables, and then, the flow of biomass as a second-stage variable. The randomness in the second-stage is included with the inclusion of the stochastic parameters: (1) cost of moisture content, (2) cost of ash content and (3) supply capacity. The aforementioned parameters vary according to the scenario, for example, the level of moisture will be different between a scenario with wet conditions and a scenario with dry conditions. **Table 5** presents the network definitions, **Table 6** shows the parameters definitions and **Table 7** introduces the variables of the model.

$$\text{Min} \sum_{j \in J} \sum_{k \in K} l_{jk} Z_{jk} + \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} \sum_{o \in \Omega} p(o) \left[c_{ij} + c_i'(t_k, o) + c_i(\delta_k, o) \right] X_{ijk}(o) \tag{5}$$

Subject to:

$$\sum_{j \in J} \sum_{k \in K} X_{ijk}(o) \leq s_i(o) \quad \forall i \in I, o \in \Omega \tag{6}$$

$$\sum_{i \in I} g_{jk} X_{ijk}(o) \leq v_{jk} Z_{jk} \quad \forall j \in J, k \in K, o \in \Omega \tag{7}$$

$$\sum_{i \in I} \sum_{j \in J} \sum_{k \in K} g_{jk} X_{ijk}(o) \geq d \quad \forall o \in \Omega \tag{8}$$

$$\sum_{k \in K} Z_{jk} \leq 1 \quad \forall j \in J \tag{9}$$

$$X_{ijk}(o) \in R^{+} \quad \forall i \in I, j \in J, k \in K, o \in \Omega \tag{10}$$

$$Z_{jk} \in \{0, 1\} \quad \forall j \in J, k \in K \tag{11}$$

Eq. (5) refers to the objective function of the stochastic model which is the minimization of the total cost (investment costs, transportation costs and quality-related costs). Eq. (6) is a constraint

Graph element	Description
N	Set of nodes in supply chain network G(N,A)
A	Set of arcs in G(N,A)
I	Set of suppliers
J	Set of potential locations for biorefineries
T	Set of arcs from *I* to *J*

Table 5. Definitions of nodes and arcs in the network graph.

Parameter	Description
l_{jk}	Equivalent annualized investment cost for opening a biorefinery in location $j \in J$ using technology $k \in K$
$p(o)$	Probability of scenario $o \in \Omega$
c_{ij}	Unit cost charged per metric ton shipped along $(i, j) \in T$
$c_i'(t_k, o)$	Quality loss due to moisture content under scenario $o \in \Omega$ for a given t_k
$c_i(\delta_k, o)$	Quality loss due to ash content under scenario $o \in \Omega$ for a given δ_k
$s_i(o)$	Supply capacity for supplier $i \in I$ for scenario $o \in \Omega$
g_{jk}	Conversion factor for biomass supplied to biorefinery $j \in J$ applying technology $k \in K$
d	Total demand of biofuel in the network N
v_{jk}	Production capacity of biorefinery $j \in J$ including technology $k \in K$

Table 6. Definitions of parameters.

Variable	Description
$X_{ijk}(o)$	Flow along arc $(i, j) \in T$ from a supplier location to a potential location for a biorefinery under scenario $o \in \Omega$
Z_{jk}	Binary variable which takes the value 1 if $j \in J$ is used as a biorefinery utilizing technology $k \in K$, and 0 otherwise

Table 7. Definitions of variables.

for the supply capacity, Eq. (7) constraints the biorefinery production capacity and Eq. (8) assures the demand satisfaction of the local market. Eq. (9) selects one technology for every open biorefinery, Eq. (10) is non-negative constraints and Eq. (11) is binary constraints.

3.3. Case study and results

Castillo-Villar et al. [16] solved a case study in the state of Tennessee to test the proposed model. The state of Tennessee has 94 counties that were considered as suppliers in the model, 31 counties were considered as potential locations for biorefineries. The biomass utilized in the model was switch-grass, all the quality information introduced in the model was derived from the experiment of Boyer et al. [17]. Three different triangular distributions (one for moisture and one for ash) were created according to number of groups included in the particle size. **Tables 8** and **9** display the parameters for those distributions.

The available biomass (19,482,102.51 dry tons) for biofuel production was obtained from the U.S. Billion-ton database, and Eq. (12) is utilized to calculate the weight of the biomass in its natural state (i.e., before drying the feedstock).

$$s_i = s_{idry}/(1 - e_i) \tag{12}$$

The scenarios were created from historical data, 11 scenarios were generated utilizing the years 2004–2014. Every region in the state of Tennessee was linked to the closest climate station to

Moisture content (%)	*at*	*t*	*bt*
Distribution 1	26	27	29
Distribution 2	17	19	20
Distribution 3	16	18	23

Table 8. The parameters of triangular distribution for the moisture content.

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

Table 9. The parameters of triangular distribution for ash content.

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

Table 10. Problem definitions.

Problem	Cost (in $millions)					
	Fixed	Transport	Moisture	Ash	Variable	Total
1	1202	41	28	27	96	1298
2	1202	41	28	37	106	1307
3	1202	46	22	24	92	1294
4	1202	46	22	31	99	1300
5	1202	45	22	23	91	1293
6	1202	45	22	29	96	1298

Table 11. Summary of experimental results.

gather information about the precipitation levels in the region. If the precipitation level of certain region is above the precipitation mean of the state, then, the region is classified as wet; otherwise it is classified as dry. A random number e_i was generated according to the classification of every region in order to calculate the moisture content. The ash content is not

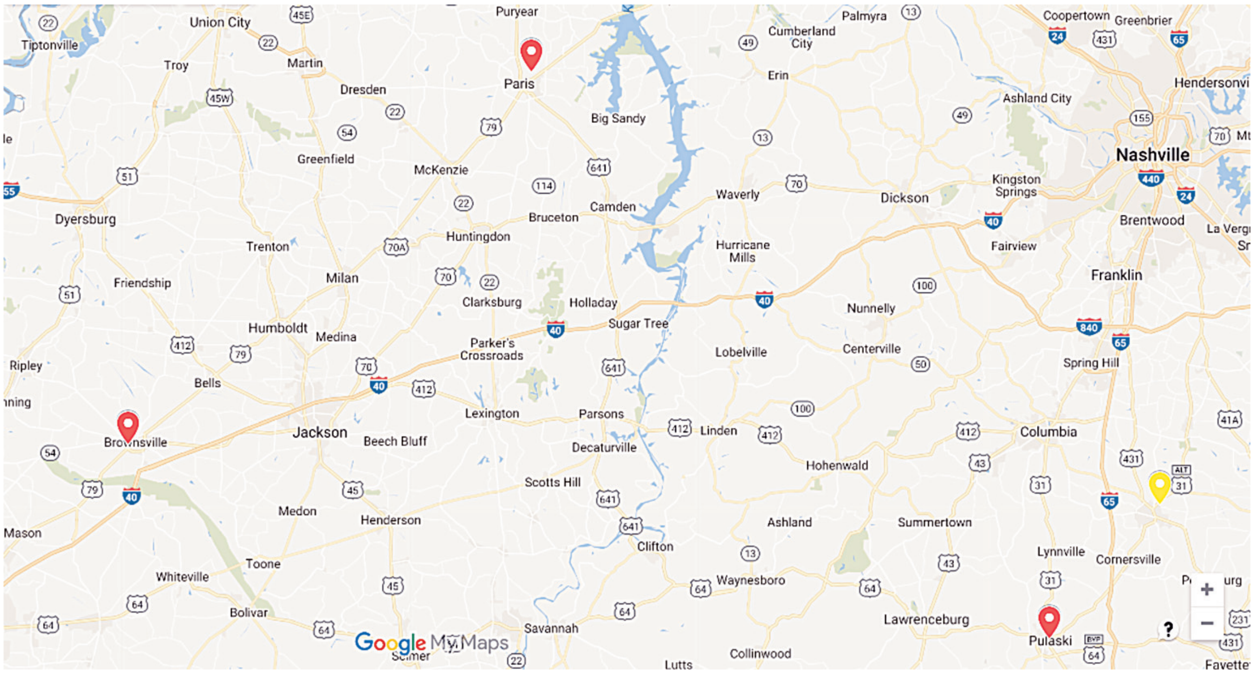


Figure 9. Solution to problem 2, including quality-related costs.

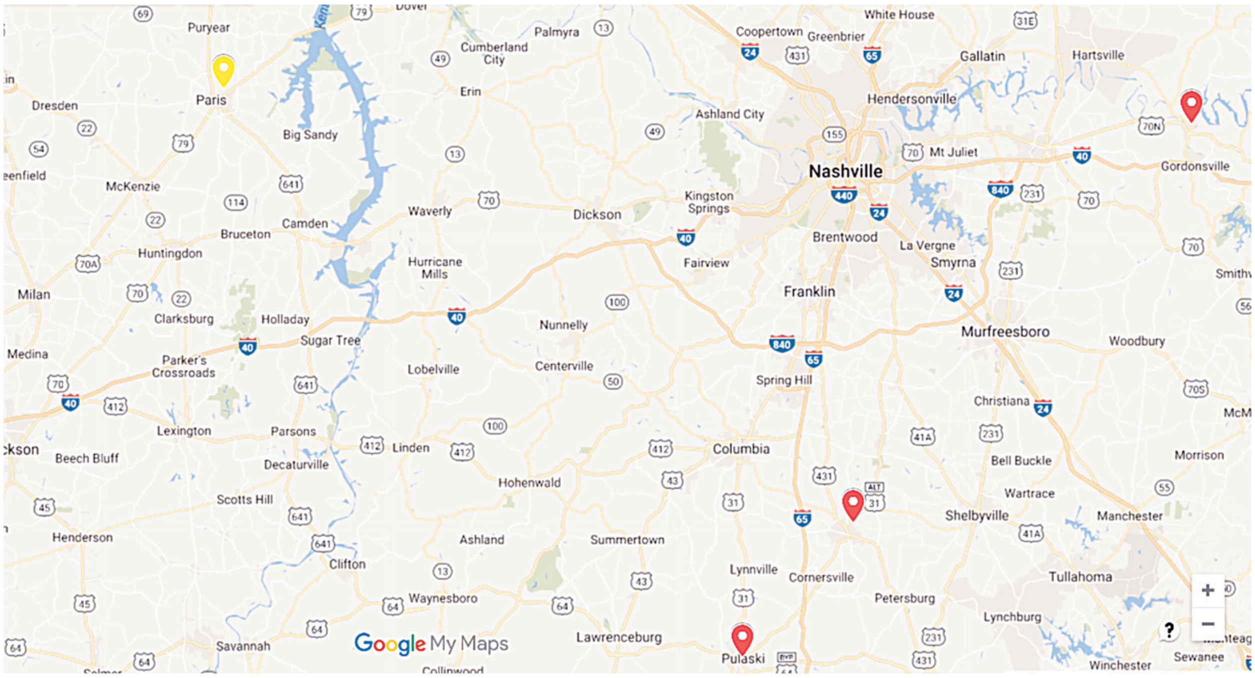


Figure 10. Solution to problem 2, without quality-related costs.

associated with the precipitation level; however, it was also contemplated in the cost calculation for the set of problems shown in **Table 10**.

The study was solved with GUROBI 6.0.0. The experiments were completed in a computer with Intel (R) Core(TM) i7-2600 U CPU @ 3.40 GHz; and 16.00 GB of RAM. The results for every problem described in **Table 10** are shown in **Table 11**. On an average, the quality related costs are about $52.5 million annually which represents a significant amount of money for an investment. **Figures 9** and **10** shows a different solution for problem 2.

Figure 9 shows the biorefineries locations for the problem 2 (high moisture, high ash) with the inclusion of the quality related costs. **Figure 10** presents the solution for problem 2 without the inclusion of the quality-related costs. The red pin indicates the location of a large capacity plant and the yellow pin shows the location of a smaller capacity biorefinery. It can be noticed that both solutions differ in the position of the pins.

4. Conclusions

This chapter presents a novel approach to incorporate physical and chemical properties of the biomass, which affect the design and implementation of the supply chain for production and distribution of biofuels. The PCA is a tool that allows the decision maker to use the variance plus additional information on how the variables interact with each other (covariance) to infer the effect of the factors included with the analysis. Moreover, PCA has the capability to detect the magnitude of the factor's effect over the variables under analysis. The factors with significant impact on the process can be incorporated into the mathematical models. The stochastic model presented in this chapter shows the impact of moisture and ash content in the production and distribution of biofuel. The results show a difference in the overall cost function, but also a different optimal design of the supply chain. Hence, ignoring the quality-related properties might lead to the cost underestimations.

Biomass has other features that impact the logistic system. Investigating new factors such as dry matter loss and integrating these factors into the modeling and algorithmic development is a fruitful future research line.

Acknowledgements

This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2014-38502-22598 and by the Hispanic Serving Institutions Education Grants Program no. 2015-38422-24064 from the USDA National Institute of Food and Agriculture.

Author details

Mario Aboytes-Ojeda and Krystel K. Castillo-Villar*

*Address all correspondence to: krystel.castillo@utsa.edu

The Texas Sustainable Energy Research Institute, The University of Texas at San Antonio, San Antonio, TX, USA

References

[1] Energy Independence. "Security Act (EISA)". In: 110th Congress, 1st session. 2007. pp. 110–140

[2] Lee RA, Jean-Michel L. From first-to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity. Animal Frontiers. 2013;**3**(2):6-11

[3] Aboytes-Ojeda M et al. A principal component analysis in switchgrass chemical composition. Energies. 2016;**9**(11):913

[4] Jacobson JJ et al. Techno-economic analysis of a biomass depot. Tech. Rep. Idaho Falls, ID, USA: Idaho National Laboratory (INL); 2014

[5] Wiselogel AE et al. Compositional changes during storage of large round switchgrass bales. Bioresource Technology. 1996;**56**(1):103-109

[6] Shinners KJ et al. Harvest and storage of two perennial grasses as biomass feedstocks. Transactions of the ASABE. 2010;**53**(2):359-370

[7] Khanchi A et al. Characteristics and compositional change in round and square switchgrass bales stored in south Central Oklahoma. Biomass and Bioenergy. 2013;**58**:117-127

[8] Kline LM et al. Investigating the impact of biomass quality on near-infrared models for switchgrass feedstocks. Bioengineering. 2016;**3**:1-22

[9] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems. 1987;**2**(1–3):37-52

[10] Shlens J. A tutorial on principal component analysis. In: arXiv preprint arXiv:1404.1100; 2014

[11] Humbird D et al. Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: Dilute-acid pretreatment and enzymatic hydrolysis of corn stover. Tech. Rep. Golden, CO: National Renewable Energy Laboratory (NREL); 2011

[12] Cundiff JS, Dias N, Sherali HD. A linear programming approach for designing a herbaceous biomass delivery system. Bioresource Technology. 1997;**59**(1):47-55

[13] Kim J et al. Design of biomass processing network for biofuel production using an MILP model. Biomass and Bioenergy. 2011;**35**(2):853-871

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

[15] Marufuzzaman M, Eksioglu SD, Huang YE. Two-stage stochastic programming supply chain model for biodiesel production via wastewater treatment. Computers & Operations Research. 2014;**49**:1-17

[16] Castillo-Villar KK, Eksioglu S, Taherkhorsandi M. Integrating biomass quality variability in stochastic supply chain modeling and optimization for large-scale biofuel production. Journal of Cleaner Production. 2017;**149**:904-918

[17] Boyer CN et al. Impact of an innovated storage technology on the quality of preprocessed switchgrass bales. Studies. 2016;**5**(12):13

IntechOpen