A Principal Component Analysis in Switchgrass Chemical Composition

Mario Aboytes-Ojeda [1], Krystel K. Castillo-Villar [1],*, Tun-hsiang E. Yu [2], Christopher N. Boyer [2], Burton C. English [2], James A. Larson [2], Lindsey M. Kline [3] and Nicole Labbé [3]

[1] Mechanical Engineering Department, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA; mario.aboytes@gmail.com

[2] Department of Agricultural and Resource Economics, University of Tennessee, Knoxville, TN 37996, USA; tyu1@utk.edu (T.-h.E.Y.); cboyer3@tennessee.edu (C.N.B.); benglish@utk.edu (B.C.E.); jlarson2@utk.edu (J.A.L.)

[3] Center for Renewable Carbon, University of Tennessee, Knoxville, TN 37996, USA; lmccull4@utk.edu (L.M.K.); nlabbe@utk.edu (N.L.)

* Correspondence: krystel.castillo@utsa.edu; Tel.: +1-210-458-8746

Abstract: In recent years, bioenergy has become a promising renewable energy source that can potentially reduce the greenhouse emissions and generate economic growth in rural areas. Gaining understanding and controlling biomass chemical composition contributes to an efficient biofuel generation. This paper presents a principal component analysis (PCA) that shows the influence and relevance of selected controllable factors over the chemical composition of switchgrass and, therefore, in the generation of biofuels. The study introduces the following factors: (1) storage days; (2) particle size; (3) wrap type; and (4) weight of the bale. Results show that all the aforementioned factors have an influence in the chemical composition. The number of days that bales have been stored was the most significant factor regarding changes in chemical components due to its effect over principal components 1 and 2 (PC1 and PC2, approximately 80% of the total variance). The storage days are followed by the particle size, the weight of the bale and the type of wrap utilized to enclose the bale. An increment in the number of days (from 75–150 days to 225 days) in storage decreases the percentage of carbohydrates by −1.03% while content of ash increases by 6.56%.

Keywords: lignocellulosic biomass; principal component analysis; statistical hypothesis; bioenergy; switchgrass

---

1. Background

Biofuel production is a promising alternative source of energy which provides advantages such as displacement of fossil-based fuels, reduction of greenhouse gas emissions, among others [1]. Biofuels can be classified according to the type of biomass utilized as follows: (1) first-generation biofuels are produced from organic sources commonly used as food for human consumption (e.g., sorghum, corn, among others); (2) second-generation biofuels are obtained from a wide range of non-food sources such as agricultural and forest residues, energy crops, and municipal solid waste; and (3) third-generation biofuels commonly refer to biofuels produced from algae [2]. This paper focuses on second-generation biofuels using a type of energy crop (i.e., switchgrass). The production of second-generation biofuels involves activities like harvesting, extracting, transporting and feedstock handling that can be enhanced to improve the biofuel supply chain performance [3,4]. Feedstock logistics play a relevant role in the optimization of the whole supply chain. Identifying factors in logistics operations that affect feedstock properties and its quality helps to improve the production of reliable biofuels.

Biomass refers to the organic feedstock from living or recently living beings, which can be used for energy production. Lignocellulosic biomass (LCB) is a plant-derived feedstock, mainly composed of cellulose, hemicellulose and lignin. Corn stover, wheat straw and switchgrass are examples of LCB with different properties such as the percentage of cellulose, hemicellulose and lignin in their chemical composition. LCB composition is directly related with the output in the energy conversion process; for example, a high percentage of cellulose leads to higher levels of biofuel production because cellulose is the raw material for biofuel production with technologies based on sugar degradation [5,6]. In supply chain operations, increasing our understanding of factors that affect the biomass composition (e.g., storage days, indoor/outdoor storage, among others) can lead to logistics improvements that maintain best operational conditions.

Densification of biomass is a common example of a supply chain operation that consists in changing the feedstock natural format by agglomerating the raw material into smaller structures. The densification process improves handling, storing, and transportation in the distribution of biomass [7,8]. Traditional densification systems include operations like baling, pelletization, extrusion and briquetting.

Densification of biomass and logistical operations involved affect quality attributes of biomass such as moisture content, unit density, bulk density, durability index and calorific value that are important for a reliable supply to biorefineries [9]. Also, densification process has shown influence in the chemical composition of biomass [10–12], which is in turn related with the aforementioned quality attributes [11,13] and with the efficiency of the conversion process. Thus, reducing the variance of biomass chemical composition can lead to high quality densified biomass such as switchgrass, corn stover, and miscanthus, among others.

Switchgrass is a perennial warm-season grass that is utilized as a lignocellulosic feedstock for biofuel production and several studies regarding its properties have been made [14–17]. Harvesting periods, pre-treatments, storage conditions and handling operations seem to affect the switchgrass chemical composition. Kline et al. [8] showed that harvesting periods affect the feedstock composition even more than soil type.

Researchers in the field have studied baling effects in feedstock properties. For example, an analysis of switchgrass composition in large round bales, stored in outdoor conditions with no protection for two different periods of 26 weeks after harvest, was performed by Wiselogel et al. [18]. Hemicellulose concentration was lower after the storage time. Moreover, ash was higher and xylan kept the same content in both periods. In the case of lignin content, the percentage was not the same in the different time intervals. The first period showed higher concentration of lignin, whereas the second period presented lower lignin content.

Another switchgrass analysis was made over bales stored in outdoor environment with different wrap material types by Shinners et al. [19]. The storage time was 311 days and after that period, cellulose concentration increased, reaching the highest value with net cover and the concentration of hemicellulose and ash also increased after the stored period, but no effect due to wrap material was observed.

Khanchi et al. [20] compared the influence of bale geometry over cellulose, hemicellulose and lignin. Two types of switchgrass bales, large round and large rectangular were analyzed after a storage period of 180 days in outdoors. The authors performed the experiment twice, each of them in a different year and both experiments showed higher cellulose content after storage time in rectangular bales. For round bales, the experiment in the first year showed a lower value for cellulose after storage while in the second year, the result was the opposite. Hemicellulose and lignin presented the same behavior in round and rectangular bales for both experiments.

In the works of Shinners et al. [19] and Khanchi et al. [20], the composition was analyzed at the beginning and at the end of the storage period. As a result, changes occurring throughout the storage period were not recorded.

Boyer et al. [21] introduced a study to show the effects of three different factors: (1) particle size, (2) wrap type and (3) days in storage in the chemical contents of switchgrass (cellulose, hemicellulose, lignin, ash and extractives percentages). Cellulose, hemicellulose, lignin, ash and extractives content changed within the groups of particle size, wrap material and days in storage. Thus, all factors showed an influence on the feedstock composition. The model utilized for this study was a univariate analysis for every single chemical component without considering relationship with other components.

This paper significantly extends previous works by:

(1)　Proposing a multivariate analysis (in this case a PCA) [22–25], with additional covariance/correlation information, to identify effects in the switchgrass composition due to selected controllable factors.
(2)　Studying the weight of the bale as a controllable factor in the chemical composition analysis (this factor has not been included in previous analysis).
(3)　Presenting a methodology to depict the contribution of factors such as storage days, particle size, wrap type and weight of the bale to the variance in the chemical composition.

Results from this statistical analysis serve to identify best practices within the most significant factors, and therefore, they can help determining conditions to enhance the quality of feedstock and, subsequently, lead to the design of robust biofuel supply chains.

The paper is structured as follows. In Section 2, methodology for PCA is presented. In Section 3, results of PCA are shown. Lastly, Sections 4–6 present the conclusions, managerial insights and future work, respectively.

2. Methodology

In 2012, an experiment to study physical and chemical properties of switchgrass was performed at the Biomass Innovation Park in Vonore Tennessee. The feedstock (Alamo Switchgrass) was harvested and baled in a rectangular shape (1.2 × 0.9 × 2.4 m) with technology New Holland. A large square baler without a cutter, New Holland BB9080 (New Holland Agriculture, New Holland, PA USA), was utilized to process the LCB at the beginning of January 2012. After being baled, the feedstock was re-located to a covered location before pre-processing, and then, on February 2012 the biomass was unpacked and grinded with a Vermeer TG5000 tub grinder (Vermeer Corporation, Pella, IA, USA). Grinded switchgrass was sampled for moisture content and chemical composition; chemical composition was determined with near-infrared (NIR) technology. After sampling, the biomass was baled one more time with a BT3 industrial baler (TLA Bale Tech LLC, South Orange, NJ, USA) in order to produce round bales (1.2 m of diameter × 1.5 m of width). The bales were moved to storage in order to begin with the experiment.

The experiment was set up with a split-split plot design. Three controllable factors were included in the experiment: (1) storage days; (2) particle size; and (3) wrap type. Storage days had three groups under study, particle size had also three groups and wrap type had two groups. The data generated was utilized in Yu et al. and Boyer et al. [14,21]. This data was also utilized to conduct the analysis presented here. This data met all necessary conditions (normality, homogeneity and heteroscedacity); thus, no further pre-processing was needed for conducting our analysis.

Detailed information related to where the biomass was grown, harvested, baled, stored, among others important features, can be found in the work of Yu et al. and Boyer et al. [14,21]. Also, additional information about the infrared methodology used to determine the chemical composition of the feedstock can be found in the study of Kline et al. [26] where a PCA was applied over the NIR spectra data, and then, information derived from PCA was used to build a partial least square (PLS) regression model to predict the chemical composition in the bales. It is worth noting that the chemical composition data generated from Kline et al. [26], was utilized for the multivariate analysis conducted in this paper to investigate four logistics (harvesting and storing) factors.

Table 1 shows the factors and variables defined for this analysis. The wrap type refers to: (1) net mesh (excluding the two ends of the round bale) and net; (2) high tensile strength film wrapping the whole round bale (net and film). The particle size was categorized in three levels, according to particle size of baled switchgrass: PS1 (243.84 cm), PS2 (7.62 cm) and PS3 (1.27–1.91 cm). Storage days were defined in three levels: 75 days, 150 days and 225 days of storage. The bale weight was included in this study by defining two levels: low (957.65–1715.20 LB) and high (1715.21–2455.10 LB). The bale weight affects the material handling processes, storage conditions and storage facilities, among others.

Table 1. Factors and variables for principal component analysis (PCA) of Switchgrass.

Factors	No. Groups within Factors	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

The variables in this analysis were: (1) cellulose—a glucose polymer linked by glycosidic bonds; (2) hemicellulose—a branched polymer of carbon sugars; (3) lignin—a structural constituent of plants, which consists of an aromatic system composed of phenyl proposal units; (4) ash—inorganic residue remaining after combustion at 550–600 °C; and (5) extractives—non-structural components that can be extracted with neutral solvents and include, but are not limited to, free sugars, proteins, chlorophyll, and waxes. The methodology to perform the multivariate analysis in this work, with factors and variable defined above, is summarized in five steps according to Figure 1:
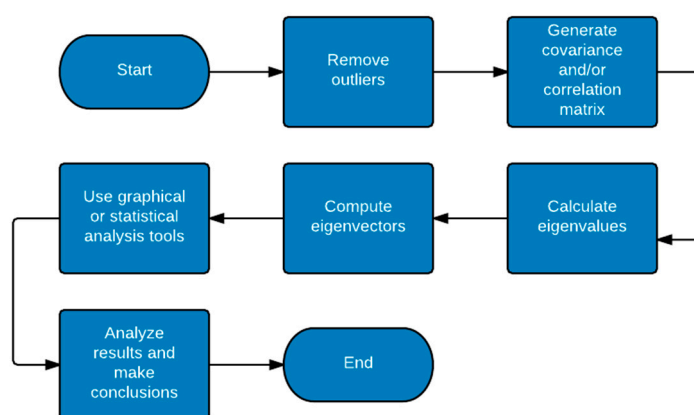
Figure 1. Flow Chart for principal component analysis (PCA).

2.1. Principal Component Analysis

In the first step of this work, no potential outliers were found and all assumptions were met (Yu et al. and Boyer et al. [14,21]). Thus, the covariance/correlation matrix was generated (refer to Table 2). The next step in the analysis was to perform an eigendecomposition, which involves computing eigenvalues and eigenvectors. The function prcomp ( ), included in the core package {stats} of R software environment [27], has been utilized to make all the calculations related with the PCA. A Bartlett's test of sphericity [28,29] was performed to determine the factorability of the correlation matrix. This test was also performed in R with the function bart_sphere ( ), included in the package {REdaS}. The resulting *p*-value was 2.22×10^{-16}; therefore, the correlation matrix did not come from a population in which the correlation matrix is an identity matrix and the correlation information for this analysis was shown to be significant.

Table 2. Covariance matrix.

Component	Cellulose	Hemicellulose	Lignin	Ash	Extractives
Cellulose	1.3442	−0.1487	−0.1298	−0.3246	−0.6855
Hemicellulose	−0.1487	1.3656	−0.4671	−0.0165	−0.3769
Lignin	−0.1298	−0.4671	0.5410	0.1978	0.0645
Ash	−0.3246	−0.0165	0.1978	0.3340	0.2263
Extractives	−0.6855	−0.3769	0.0645	0.2263	1.1001

For the analysis of eigenvalues in the covariance matrix, it is useful to know the amount of variance contained in every principal component (PC). Principal components (PCs) considered in this analysis should have the majority of the total variance (main PCs). Then, the controllable factors should be related with the main PCs in order to determine those factors with the biggest influence in the chemical composition of the feedstock under study.

According to Table 3, the first two principal components represented approximately almost 80% of the total variance. Besides the eigenanalysis, Figure 2 shows the relation between principal components and eigenvalues (variance of each PC). A common rule of thumb to determine the main PCs is that the principal components with eigenvalues above one should be included in the analysis.

Table 3. Eigenanalysis of the covariance matrix. PC: principal component.

PC	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%
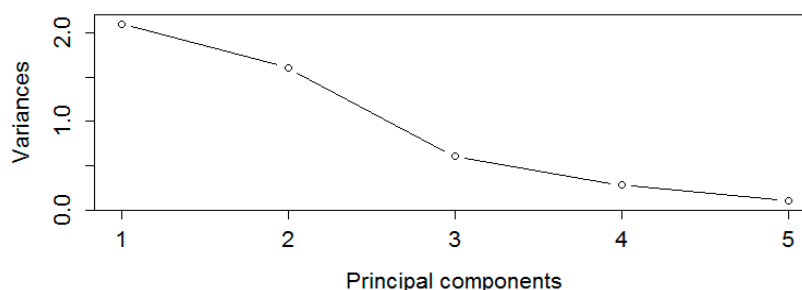
Figure 2. Scree plot for PCA.

According to Figure 2 and Table 3, PC1 and PC2 were included in this analysis (main PCs) since their eigenvalues were above one and the amount of variance was up to approximately 80%. Eigenvalues are also needed to determine the eigenvectors, that is, the orthogonal directions where largest variance is present. In Table 4, the eigenvectors are shown. The higher the eigenvector magnitude, the more correlation between the original variable and the principal component; they are either directly or inversely proportional (depending on their sign).

Table 4. Eigenvectors in principal components.

PC	1	2	3	4	5
Cellulose	−0.637	0.476	0.300	−0.519	−0.091
Hemicellulose	−0.330	−0.827	−0.048	−0.325	−0.316
Lignin	0.207	0.291	−0.655	−0.257	−0.615
Ash	0.224	−0.070	−0.324	−0.640	0.656
Extractives	0.627	−0.023	0.611	−0.386	−0.290

The more correlated variables within each principal component were:

(1) PC1—cellulose and extractives (C&E).
(2) PC2—hemicellulose (HEM).
(3) PC3—lignin and extractives (L&E).
(4) PC4—cellulose and ash (C&A).
(5) PC5—lignin and ash (L&A).

An improvement in the value of PC1 (C&E) resulted in an increment in the value of extractives and a decrement in the value of cellulose, whereas a reduction in this principal component implied an improvement in the cellulose value and a decrease in the value of extractives, as shown in Table 4. Similarly, in PC2 (HEM), an increase in the value of this component represented a reduction in the value of hemicellulose and vice versa.

With the eigenvector information, original data can be transformed into a set of multivariate data points located in the principal components (scores). The set of scores can be analyzed in a visual manner with exploratory graphs, in order to detect segregation between groups for any controllable factor selected. Furthermore, if there is no distinguishable segregation in the exploratory graph, a statistical test can be implemented to determine the differences in their means, and, therefore, the effect of the corresponding factor. In this work, an exploratory graph know as bi-plot chart was utilized. Then, if no visual segregation between the groups was detected, a pairwise *t*-test with adjusted *p*-value, was implemented for the statistical analysis. The function autoplot ( ) in the package {ggfortify} and ordiellipse ( ) in the package {vegan} (R software environment [27]) were executed for this analysis. The function pairwise.t.test ( ) included in the package {stats}, also in R, was run in the comparison analysis. The next section introduces the statistical test for pairwise comparison.

2.2. T-test in Principal Components

In some instances of the analysis, it is not possible to visually identify segregation within the groups of a factor, and therefore, a statistical test is required to determine an effect over the chemical composition of the feedstock [30]. In this work, pairwise *t*-test was utilized to make comparisons between the groups in order to conclude an influence over PCs due to the difference in their score means. A mean difference within the PCs corresponds to a difference in the chemical composition (i.e., the level of cellulose, hemicellulose, lignin, ash and extractives).

The test for this difference considered unknown but equal population variances, known sample means and not equal sample variances. The test was performed using the *t*-distribution and the corresponding estimator is defined by the following expression:

$$t = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{1}$$

In the definition of parameter *t*, $\overline{x_1}$ represents the sample mean of the first group under study whereas $\overline{x_2}$ corresponds to the sample mean of the second one. Similarly, μ_1 represents the population mean for the first group and μ_2 to the sample mean for the second one. In the denominator, s_p is the pooled estimator in the case where the sample variances are not equal. Also, n_1 and n_2 represent the corresponding number of samples for the first and second group in the analysis. The following function describes this pooling effect:

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

In Equation (2), s_1 represents the standard deviation within the first group and s_2 within the second group. With this approach, it was possible to statistically test every factor through its groups on every principal component, and, therefore, it was possible to see the most important factors in terms

of variance. All the results for the pairwise statistical tests of these interactions were displayed in a matrix format. The main findings from this matrix are discussed in Section 3.2.

3. Results

3.1. Bi-Plot Chart Analysis in PC1 and PC2

The objective in this exploratory analysis is to visually identify segregations within the groups, for every factor to identify an effect in the chemical composition. Also, the bi-plot chart shows the eigenvectors to determine the relation between PCs, original variables and the groups within the factors.

The exploratory analysis of factors in the plane PC1-PC2 was performed through the bi-plot analysis [31]. Score points are plotted in an *XY*-plane (C&E—HEM plane). Then, they are differentiated in the plot through colors and symbols, according to their related factor level or group in order to identify an emerging pattern. Figure 3 shows no identifiable clusters between net wrap and net and film wrap. Hence, there is no apparent effect of this factor since the dispersion for their scores seems to be very similar.
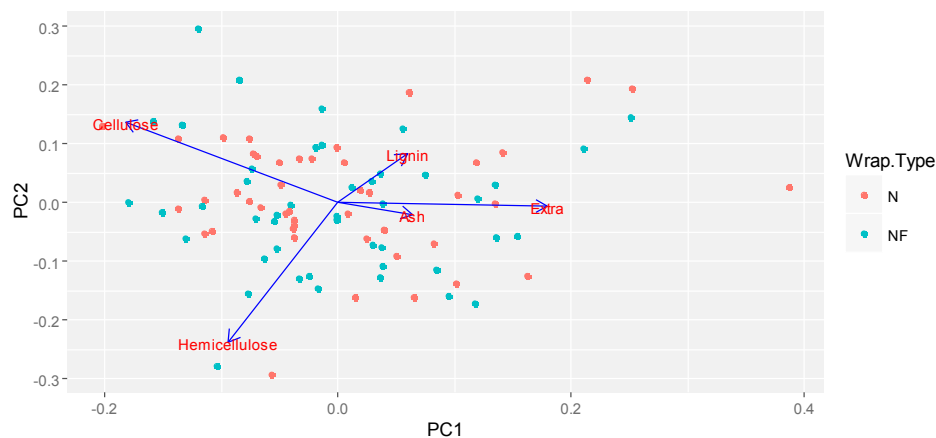

Figure 3. Bi-plot chart for analysis of wrap type.

The next factor in the analysis is bale weight. Figure 4 presents no difference between the groups of bale weight regarding the chemical composition of switchgrass, since there is no visual distinction of any clustering within the factor. Similarly, Figure 5 shows the particle size scores and there is no distinction of clustering either, so there is no effect linked to changes in the composition.
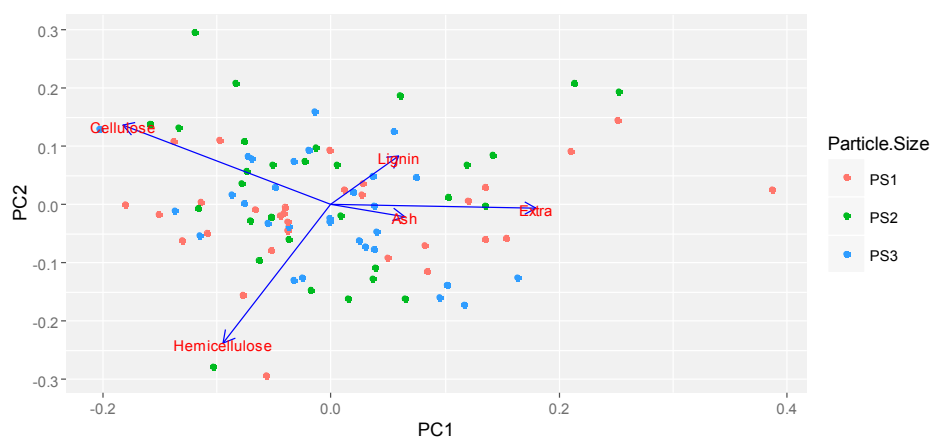

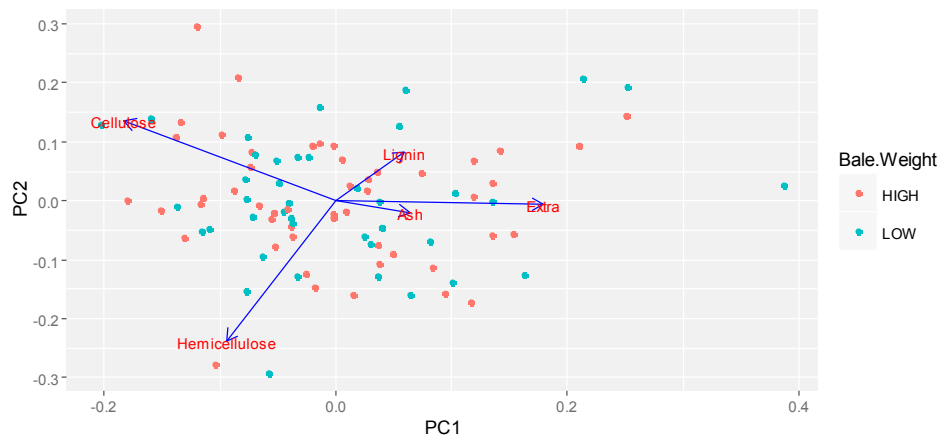**Figure 4.** Bi-plot chart for analysis of particle size.

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

Figure 6 displays an apparent segregation between the groups related to storage days and also shows that the groups related to 75 and 150 storage days are located in the upper part of the plot while the group related to 225 of storage days is located at the lower region of this graph. To improve the exploratory analysis of the effect due to storage days, Figure 7 shows only two groups (two ellipses) related to storage days. One group has been created by joining the original groups associated with 75 and 150 storage days (<150 days), the second group considers only the data for 225 storage days.
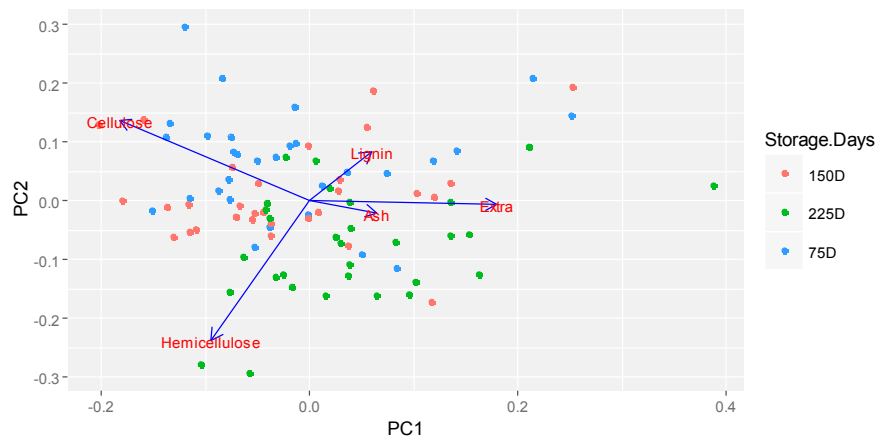

Figure 6. Bi-plot chart for analysis of storage days.
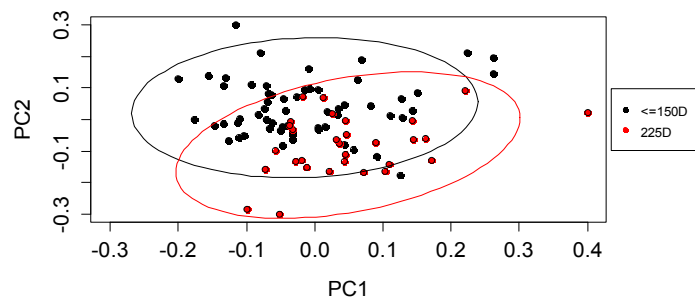

Figure 7. Score plot analysis dispersion according to the storage days (re-grouped).

The score plot shows a horizontal displacement in the ellipses corresponding to a segregation effect in PC1 which contributes with 34% of variance. A vertical displacement is also distinguishable within the ellipses; that segregation corresponds to the influence over PC2 with 45% of the total

variation. Thus, the variance percentage associated with the number of storage days can be estimated via visual inspection since they affect the first and second component (C&E and HEM); the number of days in storage is related with the 79% of the variance.

Figure 6 also shows the relation between original and PCs variables, as well as the relation with the groups within storage days. According to the bi-plot chart, the direction where cellulose increases points to the quadrant located at the upper left part of the chart, with negative values for PC1 and positive values for PC2. This quadrant is dominated with blue points, and therefore, there is a visual relation between high levels of cellulose and 75 days of storage.

The hemicellulose has a positive direction towards negative PC1 and PC2 values, where most of the points related with 225 days of storage are located, then, an increase in the percentage of hemicellulose was observed, as the number of days in storage days increases as well. The relation of cellulose-storage days and hemicellulose-storage days are opposite. Table 5 shows that data analysis in original variables is consistent with the exploratory analysis in the bi-plot chart.

Table 5. Means, deltas and %deltas in chemical composition due to storage days.

Index	Factor	Group	Cellulose	Hemicellulose	Lignin	Ash	Extra	Carbs
(1)	Storage days	$\leq$150 D	40.58	26.58	22.61	2.09	5.23	67.17
(2)	Storage days	225 D	39.13	27.35	22.42	2.23	5.53	66.48
(3)	Delta	= (2) − (1)	−1.46	0.77	−0.18	0.14	0.30	−0.69
(4)	%Delta	= (3)/(1)	−3.59%	2.89%	−0.80%	6.56%	5.79%	−1.03%

Finding factors that influence the main principal components is necessary to shed some light on the changes in chemical composition of switchgrass, and therefore, to improve the feedstock quality by reducing its variance.

Since all the factors have been analyzed through exploratory techniques, and there is still a lack of association between controllable factors and PCs, the next step is to implement statistical analysis over the complete set of principal components. The results of pairwise *t*-tests can provide validity to those insights revealed by the bi-plot chart.

3.2. Pairwise T-test

Several *t*-tests of pairwise comparisons of the controllable factors considered in this work were performed. Forty tests were run in total: 5 comparisons for wrap type, 15 for particle size, 15 for storage days and 5 for bale weight. A correction for multiple comparisons was applied in order to deal with error type I [32]. Figure 8 shows the results for every test made. Each column in Figure 8 has a principal component and each row has a controllable factor in the analysis.

The storage days are the only factor that has statistical significance over scores in the first and second principal component according to *t*-test results. Recall that the PCA shows that PC1 and PC2 contain almost 80% of the total variance in the chemical composition. In fact, storage days are significant in every principal component (as shown in Figure 8), and therefore, is the only factor that will affect the total variance in the chemical composition.

The particle size has influence over the chemical composition according to the results of statistical *t*-test performed in PC3 and PC4 (which account for 18.83% of the variance). The bale weight shows an effect on switchgrass composition in PC3 and PC4 (which account for 18.83% of the variance). Finally, wrap type have an effect in PC3 (which account for 12.90% of the variance). The pairwise *t*-tests results confirmed the insights from the exploratory graphs.
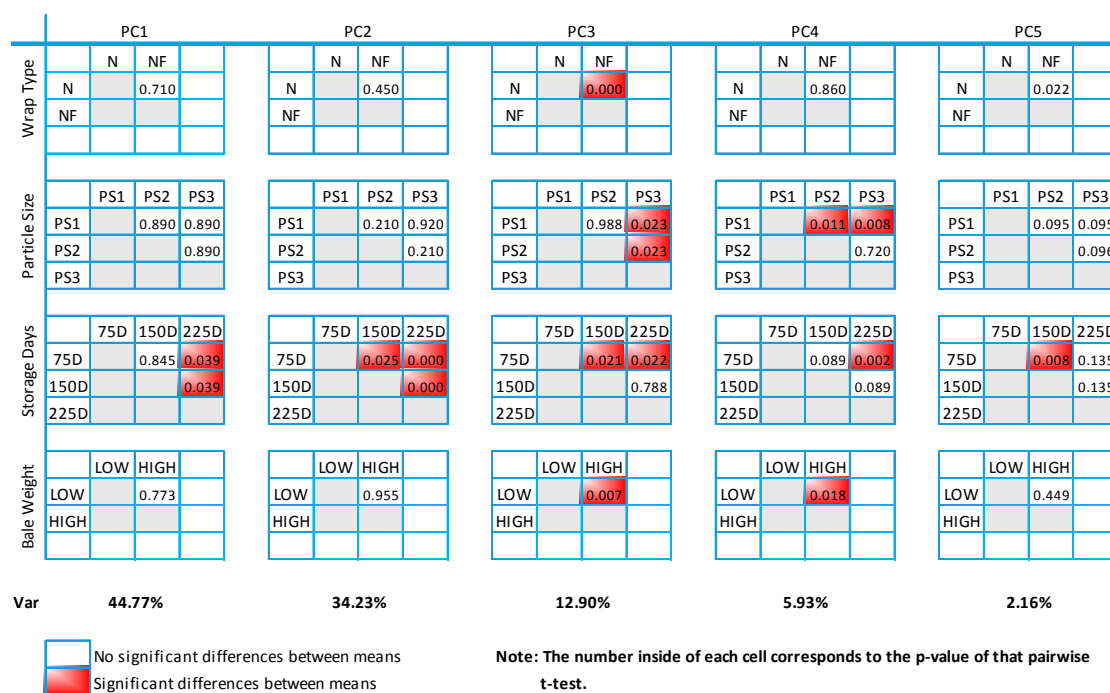
Wrap Type

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

Particle Size

	PC1 PS2 / PS3	PC2 PS2 / PS3	PC3 PS2 / PS3	PC4 PS2 / PS3	PC5 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

Storage Days

	PC1 150D / 225D	PC2 150D / 225D	PC3 150D / 225D	PC4 150D / 225D	PC5 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

Bale Weight

	PC1 HIGH	PC2 HIGH	PC3 HIGH	PC4 HIGH	PC5 HIGH
LOW	0.773	0.955	0.007	0.018	0.449

| 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 every factor considered in the analysis.

4. Conclusions

Particle size, bale weight and wrap type have an influence on the chemical composition of switchgrass; however, the PCA results show that the aforementioned effects are small in comparison to the effect of storage days. The number of days in storage affects the variance within the chemical composition. Storage days represents a controllable factor that can be fixed to a value that reduces the variance in switchgrass composition so that the quality of the feedstock can be improved.

According to Figure 8 and Table 5, it can be concluded that the increment in storage days ($\leq$150–225 D) has a decrement effect on the carbohydrates (i.e., -1.03% in cellulose and hemicellulose). Also, the aforementioned change in the number of days in storage, increases the percentage of ash in the switchgrass composition in 6.56%. The design of a supply chain should implement strategies to reduce the numbers of days that feedstock will remain in storage due the reduction in carbohydrates and increase in ash content. Lowering the percentage of ash will increase the likelihood of meeting the target of 1% when using thermochemical technology. Thus, based on the results, storing the switchgrass for less than 150 days in an inventory is recommended.

5. Managerial Insights

The results of this work can be used to design and improve switchgrass logistics. Storage days are associated with most of the variance, followed by particle size, bale weight and wrap material. The supply of biomass should be mostly designed and implemented according to the storage days with best performance regarding feedstock quality and then, a just-in-time philosophy can lead to get the right quantity of feedstock at the right place and the right time. To achieve this goal, prediction models are needed to generate valuable information.

Weight of the bale and particle size are related to feedstock composition. Therefore, this information becomes relevant to design the bale and all the logistics (e.g., handling and transportation).

Since storage days, particle size, weight of the bale and wrap material affect the switchgrass quality; statistical tools can also be used to track the quality performance of switchgrass bales and detect important deviations from specifications. This detection can lead to finding the root causes, assuring that the required biomass quality for the conversion process is met.

6. Future Work

Storage days are the most relevant factor in this study. Thus, future research should increase the knowledge in this particular factor. This work introduced three different storage periods. However, relevant information within those periods might need further investigation. Prediction models to better estimate the chemical composition in feedstock should be based in time-series analysis (e.g., moving average, exponential smoothing, regression analysis, and autoregressive integrated moving average, among others). In order to develop such models, it is recommended to gather sufficient data and adopt a multivariate approach. The insights related with storage days can be utilized to set restrictions in decision models. For instance, the bale of switchgrass cannot be stored for more than *n* days due the changes in chemical composition. Moreover, this can be reflected in the objective function of an optimization model by including the opportunity cost in a profit or cost function.

Quality assurance is also a promising area of research due to its role in the reliability and economically feasibility of bioenergy production. Development of tools such as statistical quality control within a multivariate frame is a plausible approach since biomass has several properties that are involved within its energy extraction. Assuring that these properties meet the specifications is vital to detect and prevent any issues in the conversion process.

Physical constitution of feedstock bales is also a promising line of research, the weight and the particle size can be critical in logistical decisions because they affect the chemical composition and, hence, the production yield of biofuels. Therefore, more research including factors related to physical properties in the biomass supply chain decision making and management is needed.

Author Contributions: Mario Aboytes-Ojeda and Krystel K. Castillo-Villar proposed the methodology to depict the contribution of controllable factors to the variance of the chemical composition, conducted the statistical analysis, and wrote the paper; Tun-hsiang E. Yu, Burton C. English, James A. Larson and Christopher N. Boyer conceived, designed and performed the experiments; Lindsey M. Kline and Nicole Labbé performed the chemical analysis of the stored biomass.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Antizar-Ladislao, B.; Turrion-Gomez, J.L. Second-generation biofuels and local bioenergy systems. *Biofuels Bioprod. Biorefining* **2008**, *2*, 455–469. [CrossRef]
2. Lee, R.A.; Lavoie, J.-M. From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity. *Anim. Front.* **2013**, *3*, 6–11. [CrossRef]
3. Sokhansanj, S.; Kumar, A.; Turhollow, A.F. Development and implementation of integrated biomass supply analysis and logistics model (IBSAL). *Biomass Bioenergy* **2006**, *30*, 838–847. [CrossRef]
4. Mafakheri, F.; Nasiri, F. Modeling of biomass-to-energy supply chain operations: Applications, challenges and research directions. *Energy Policy* **2014**, *67*, 116–126. [CrossRef]
5. U.S. Department of Energy. *Bioenergy Research Centers*; DOE/SC-0162; Office of Biological and Environmental Research: Washington, DC, USA, 2014.
6. Rubin, E.M. Genomics of cellulosic biofuels. *Nature* **2008**, *454*, 841–845. [CrossRef] [PubMed]
7. Shaw, M.D. Feedstock and Process Variables Influencing Biomass Densification. Master's Thesis, University of Saskatchewan, Saskatoon, SK, Canada, 2008.
8. Kline, L.; Johnson, A.; Kim, P.; Jackson, S.; Labbe, N. Monitoring switchgrass composition to optimize harvesting periods for bioenergy and value-added products. *Biomass Bioenergy* **2013**, *56*, 29–37.
9. Tumuluru, J.S.; Wright, C.T.; Kenny, K.L.; Hess, J.R. *A Review on Biomass Densification Technologies for Energy Application*; Idaho National Laboratory (INL): Idaho Falls, ID, USA, 2010.

10. Rijal, B.; Igathinathane, C.; Karki, B.; Yu, M.; Pryor, S.W. Combined effect of pelleting and pretreatment on enzymatic hydrolysis of switchgrass. *Bioresour. Technol.* **2012**, *116*, 36–41. [CrossRef] [PubMed]

11. Raveendran, K.; Ganesh, A. Heating value of biomass and biomass pyrolysis products. *Fuel* **1996**, *75*, 1715–1720. [CrossRef]

12. Jacobson, J.J.; Roni, M.S.; Lamers, P.; Cafferty, K.G. *Feedstock Supply System Design and Analysis*; Idaho National Laboratory (INL): Idaho Falls, ID, USA, 2014.

13. Yaman, S. Pyrolysis of biomass to produce fuels and chemical feedstocks. *Energy Convers. Manag.* **2004**, *45*, 651–671. [CrossRef]

14. Yu, T.E.; Larson, J.A.; English, B.C.; Boyer, C.N.; Tyler, D.D.; Castillo-Villar, K.K. Influence of particle size and packaging on storage dry matter losses for switchgrass. *Biomass Bioenergy* **2015**, *73*, 135–144. [CrossRef]

15. McLaughlin, S.B.; Kszos, L.A. Development of switchgrass (Panicum virgatum) as a bioenergy feedstock in the United States. *Biomass Bioenergy* **2005**, *28*, 515–535. [CrossRef]

16. Sanderson, M.A.; Adler, P.R.; Boateng, A.A.; Casler, M.D.; Sarath, G. Switchgrass as a biofuels feedstock in the USA. *Can. J. Plant Sci.* **2006**, *86*, 1315–1325. [CrossRef]

17. Lee, D.; Owens, V.N.; Boe, A.; Jeranyama, P. *Composition of Herbaceous Biomass Feedstocks*; North Central Sun Grant Center, South Dakota State University: Brookings, SD, USA, 2007.

18. Wiselogel, A.E.; Agblevor, F.A.; Johnson, D.K.; Deutch, S.; Fennell, J.A.; Sanderson, M.A. Compositional changes during storage of large round switchgrass bales. *Bioresour. Technol.* **1996**, *56*, 103–109. [CrossRef]

19. Shinners, K.J.; Boettcher, G.C.; Muck, R.E.; Weimer, P.J.; Casler, M.D. Harvest and storage of two perennial grasses as biomass feedstocks. *Trans. ASABE* **2010**, *53*, 359–370. [CrossRef]

20. Khanchi, A.; Jones, C.L.; Sharma, B.; Huhnke, R.L. Characteristics and compositional change in round and square switchgrass bales stored in South Central Oklahoma. *Biomass Bioenergy* **2013**, *58*, 117–127. [CrossRef]

21. Boyer, C.N.; Yu, E.T.; English, B.C.; Larson, J.A.; Labbé, N.; Kline, L.M. Impact of an innovated storage technology on the quality of preprocessed switchgrass bales. *AIMS Bioeng.* **2016**, *3*, 125–138. [CrossRef]

22. Shlens, J. A Tutorial on Principal Component Analysis. 2014. Available online: https://arxiv.org/abs/1404.1100 (accessed on 21 September 2015).

23. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [CrossRef]

24. Bro, R.; Smilde, A.K. Principal component analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [CrossRef]

25. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]

26. Kline, L.M.; Labbé, N.; Boyer, C.; Yu, T.E.; English, B.C.; Larson, J.A. Investigating the impact of biomass quality on near-infrared models for switchgrass feedstocks. *AIMS Bioeng.* **2015**, *3*, 1–22. [CrossRef]

27. The R Project for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 19 October 2015).

28. Bartlett, M. Tests of significance in factor analysis. *Br. J. Stat. Psychol.* **1950**, *3*, 77–85. [CrossRef]

29. Bartlett, M.S. The Effect of standardization on a χ^2 approximation in factor analysis. *Biometrika* **1951**, *38*, 337–344. [CrossRef]

30. Montgomery, D.C. *Introduction to Statistical Quality Control*; John Wiley & Sons: Hoboken, NJ, USA, 2007.

31. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]

32. Benjamini, Y.; Hochberg, Y.; Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300.