Multivariate data analysis applied in alkali-based pretreatment of corn stover

Bin Li, Li Ding, Huanfei Xu, Xindong Mu, Haisong Wang

A R T I C L E   I N F O

Article history:
Received 25 January 2016
Received in revised form 16 December 2016
Accepted 16 December 2016

Keywords:
Multivariate data analysis
Lignocellulosic biomass
Alkali-based pretreatment
Enzymatic saccharification

A B S T R A C T

In this paper, variables of Pulp Refining Instrument (PFI) refining assisted alkaline pretreatment and hydrolysis saccharification of corn stover were analyzed. The process parameters were characterized by multivariate data analysis methods including Principle Component Analysis and Partial Least Square (PCA and PLS) to investigate the specific relationships of primary variables in alkali-based pretreatment of biomass. In this paper, pretreatment system was multivariate and the variables were inter-related to each other. Total alkaline charge and removal rate of lignin had greatest impact on the pretreatment and enzymatic hydrolysis.

© 2016 Published by Elsevier B.V.

1. Introduction

In view of the shortage of fossil fuels and the worsening environmental problems all over the world, more researchers have focused on developing of sustainable bio-energy alternatives (Bichraoui-Drazer et al., 2015; Ayamga et al., 2015). Bio-energy was one of the sustainable energy alternatives (Ding et al., 2012; Yang et al., 2013; Von Schenck et al., 2013; Sun and Cheng, 2002). During the bio-energy conversion process, enzymatic saccharification is one of the key steps to produce fermentable sugars, which can be further converted to bio-fuels (such as bio-ethanol) (Yu et al., 2013). The lignocellulosic biomass was considered as the most promising feedstock of the second-generation bio-ethanol (Ljungqvist et al., 2005). Main components of lignocellulosic biomass were polysaccharide (cellulose and hemicellulose) and lignin (Zhu et al., 2009; Kho et al., 2015). As polysaccharide, cellulose and hemicellulose could be degraded to monosaccharide (fermentable sugars), which could be converted into bio-ethanol (Himmel et al., 2007; Luo et al., 2011).

Thus, the pretreatment was considered as the key process to break the natural barriers of lignocellulosic biomass (Segal et al., 1959; Luo and Zhu, 2011). High efficient pretreatment could reduce the subsequent operating cost of the bioenergy conversion. There were many different biomass pretreatment categories (Oh et al., 2005), such as physical pretreatment, biological pretreatment, chemical pretreatment, and physicochemical pretreatment (Liu et al., 2013). Therein, the physicochemical pretreatment was considered as one of the high efficiency pretreatment to improve the enzymatic saccharification (Agbor et al., 2011). Enzymatic saccharification was normally produced by enzymatic hydrolysis process, in which polysaccharide could be decomposed into fermentable sugar by enzyme (e.g. cellulose, glucosidase).

Many different process variables were in the physicochemical pretreatment and enzymatic hydrolysis (Miura et al., 2012). More or less, these variables were associated with each other. In consideration of the relationship between the different variables (Li et al., 2011), studies of the pretreatment conditions, mechanical treatment, composition of pretreated corn stover, enzymatic hydrolysis effect could have an important guiding significance.

When multiple variables were involved, the data could be analyzed by a comprehensive dimension reduction method, which could summarize the main aspects or information of a dataset (Lin et al., 2015). The multivariate data analysis (PCA and PLS) were especially useful for reducing high-dimension multivariate systems (Godoy et al., 2014). On account of multiple linear regressions, the orthogonal transformation could be used by PCA method to trans-
form the possibly correlated variables base set into a new base set of linearly uncorrelated variables (principal components). By PCA analysis, the variables could keep independent, and the data noise could be removed to help find the key variables for subsequent analysis. As a statistical method, PLS could reveal the relationship of principal components regression. PLS method could not just give the hyperplanes of minimum variance (response vs independent variables). PCA and PLS could simplify the complex process or issues, and reveal the essence of the variables (Li et al., 2014).

In this paper, the lignocellulosic biomass corn stover was treated by physicochemical pretreatment (PFI refining assisted alkaline pretreatment) and enzymatic hydrolysis process. Variables of the pretreatment and enzymatic hydrolysis were analyzed by PCA and PLS method to find the inter-relationships between each other in order to guide the further research of lignocellulosic biomass conversion.

2. Materials and methods

2.1. Raw material and chemicals

Corn stover (lignocellulosic biomass raw material) was harvested in the October in 2014, from Pingdu, China. Firstly, corn stover was milled by a grinder. Then, the milled samples were screeneded to collect the corn stover with the target particle size (0.18–0.85 mm). After that, the samples were stored in a sealed bag at temperature for following research. (The moisture content and composition analysis were shown in supporting information Table 1).

Sodium hydroxide was bought from Sinopharm Chemical Reagent Co., Ltd. Cellulase (Celluclast 1.5L) and β-glucosidase (Novozyme 188) were supplied by the Sigma-Aldrich China Inc. All enzymes and chemicals were used without further purification.

2.2. Alkaline pretreatment of corn stover

The alkaline pretreatment was conducted in a cooking reactor (Mode PL1-00, Xianyang TEST Equipment Co., Ltd., Xianyang, China). Four small cooking seal pots (1 L) were in the cooking reactor. For every small cooking pot, 50 g oven dried raw biomass was added with alkaline chemicals at certain ratio of liquid to solid and certain temperature for certain holding time. (The pretreatment conditions could be seen in supporting information Table 2). After alkaline pretreatment, the cooking seal pots were cooled down quickly. Then, the pretreated corn stover samples was treated to be neutral by washing with water in a 300 mesh bag. Finally, the pretreated corn stover samples were stored at 4 °C for further study.

2.3. Mechanical refining treatment

The corn stover samples with alkali-based pretreatment were experienced the mechanical refining treatment (PFI) by a refiner (model PL1-00, Xianyang TEST Equipment Co., Ltd., China). The revolution numbers (corn stover samples treated by PFI between the fixed disc and rotating disc) of the PFI treatment in this work were 4000. The oven dried weight of corn stover samples was 10 g. The corn stover solid content and water content of the pretreated samples in PFI treatment were 10% and 90%, respectively. The speed of the PFI mechanical refiner was 1400 rpm (0.24 mm refining gap).

2.4. Enzymatic hydrolysis

In the enzymatic hydrolysis process, the alkali-based pretreated biomass samples were conducted in a 25 ml bottle with sodium citrate buffer and 0.02% sodium azide (0.05 M, pH 4.8), at 50 °C for 48 h. The process was produced in a water bath shaker (90 rpm).

For 1 g dry pretreated corn stover, the dosage of β-glucosidase was 5 IU/g and the celulase was 20 FPU. Then, the supernatant liquid after enzymatic hydrolysis was collected and placed in a sealed bag at −10 °C for following experiment.

2.5. Analysis methods

National Renewable Energy Laboratory (NREL) protocol (Sluiter et al., 2008) was used to test the chemical composition of corn stover. A high performance liquid chromatography (HPLC)(Model 1200, Agilent, USA) with a Bio-Rad Aminex HPX-87H column (300 × 7.8 mm) was used to determine the enzymatic hydrolyzate at 55 °C with 0.005 M sulfuric acid.

The evaluation equations were as follows:

\[ R_{\text{solid}}(\%) = \frac{\text{weight of pretreated sample (g)}}{\text{untreated sample (g)}} \times 100\% \]

\[ R_{\text{glucan/xylan}}(\%) = \frac{(R_{\text{solid}} \times D_{\text{glucan/xylan of pretreated corn stover}})}{D_{\text{glucan of raw corn stover}} \times 100\%} \]

\[ R_{\text{delignification}}(\%) = 1 - \frac{(R_{\text{solid}} \times D_{\text{lignin of pretreated corn stover}})}{D_{\text{lignin of raw corn stover}} \times 100\%} \]

\[ E_{\text{glucan}}(\%) = \frac{(A_{\text{glucose in hydrolyzate}} \times 0.9)}{A_{\text{glucan in pretreated corn stover}} \times 100\%} \]

\[ E_{\text{xylan}}(\%) = \frac{(A_{\text{xyllose in hydrolyzate}} \times 0.88)}{A_{\text{xylan in pretreated corn stover}} \times 100\%} \]

2.6. PCA and PLS analysis

The data analysis for PCA model and PLS model was calculated by the SIMCA-P 12.0 (Soft Independent Modelling of Class Analogy, Umetrics).

3. Result and discussion

3.1. Principal component analysis of the whole process

In this paper, the multivariate data analysis included PCA and PLS methods. The PCA was known as analysis for principal component, which was designed to reduce the dimension in order to reduce many indicators into a few composite ones. To comprehensively and systematically analyze the problems, the relationship between multiple different factors should be considered. Commonly, these factors or indicators in multivariate statistical analysis were known as variable. Since each variable reflected some information of the research issue in some extent, and each variable had a certain correlation with each other, the statistics information reflected by all the variables was partly overlapped. In the multivariate analysis study, too many variables would increase the amount of calculation and the analysis complexity. As a result,
in the quantitative analysis, the PCA method would reduce the number of variables, and at the same time get more valuable information. In the PCA method, many models were established to help the researchers understand the relationship between various variables. The models would reveal the similarities and dissimilarities between the data points and recognize the abnormal value in the issue (Parthivos and Petter, 2013).

The PLS method was a mathematical optimization technology, which could find the best matching function by minimizing the square error. The systematic PLS model could be established between the independent variables and the dependent variable, which could connect the main component in the principal component analysis with the meaningful data points. The PLS model could reveal the relationship between the independent variables and the dependent variables in the multivariate analysis.

In this paper, the corn stover was pretreated by sodium hydroxide based alkaline pretreatment. The 18 key process parameters in the pretreatment and enzymatic saccharification process were analyzed by PCA and PLS methods. The 18 key parameters were as follows: the total alkaline amount calculated based on the corn stover (Alkaline%), the ratio of solid to liquid (S/L ratio), the maximum holding time of the pretreatment temperature (H time), with or without PFI treatment (PFI), the glucan content of pretreated corn stover (Glucan%), the xylan content of pretreated corn stover (Xylan%), the extractive content of corn stover (Extractive%), the lignin content of corn stover (Lignin%), the solid yield after pretreatment (Solid Yield%), the desorption of lignin after pretreatment (D-lignin%), the recovery rate of glucan (R-glucan%), the recovery rate of xylan (R-xylan%), the enzymatic hydrolysis efficiency of glucan (E-glucan), the enzymatic hydrolysis efficiency of xylan (E-xylan%), the final total yield of glucan (Y-glucan%), the final total yield of xylan (Y-xylan%), the final total sugar yield (Y-total sugar%). These 18 parameters were divided into four categories, as shown in Table 1. All the data in this pretreatment research were imported into the SIMCA-P software.

![Image](image_url)

**Table 1**

| classification of variables | Name of variable | Value range of variable |
|-----------------------------|-----------------|------------------------|
| The parameters of pretreatment conditions | Alkaline (%) | 5–13% |
| | S/L ratio | 4–8 |
| | T: C | 120–160 |
| | H time (min) | 10–50 min |
| | PFI | 0 with \%1 without |
| The compositions of pretreated corn stover | Glucan (%) | 40.1–55.5 |
| | Xylan (%) | 22.5–27.4 |
| | Extractive (%) | 3.41–9.3 |
| | Lignin (%) | 2.3–12.4 |
| Evaluation of pretreatment | Solid Yield (%) | 44.6–68.2 |
| | D-lignin (%) | 22.1–77.9 |
| | R-glucan (%) | 53.2–92.5 |
| | R-xylan (%) | 52.0–80.7 |
| The evaluation of enzymatic effect | E-glucan (%) | 51.1–85.4 |
| | E-xylan (%) | 34.5–75.1 |
| | Y-glucan (%) | 54.3–85.1 |
| | Y-xylan (%) | 36.7–58.6 |
| | Y-total sugar (%) | 47.9–74.4 |

After the original data was analyzed by SIMCA-P software, the PCA models were established, as shown in Fig. 1.

The first step of multivariate analysis was principal component analysis, which could obtain an overall graph based on the entire data set. As mentioned above, PCA method could reduce the number of the original variables for main component. The first principal component (P1) was in the direction of the biggest change of the data set, and the second principal component (P2) was in the orthogonal direction of P1. In the successive principal component, the amount of the variable of each principal component was less than the amount of the previous one. As shown in Fig. 1.a, the PCA model showed four main variables. The value of R square was 0.795; the value of Q square was 0.644. The value of R square could represent the proportion of the variable for all the original variables, and the closer value of R square was to 1, the better analysis results the PCA model could get. The value of Q square could represent the predictive ability of the PCA model, which was calculated by the cross-validation method. When the value of the square of Q was greater than 0.5, the PCA model could get a good predictive ability. As a result, the original experimental data generated a relatively credible PCA model.

As can be seen in the scatter plot figure (Fig. 1b), the relationship between the variables was reflected by the first two main component of PCA model (P1 and P2), which explained about 60% information of dataset. The farther away the variable point was from the origin point of coordinates the greater impact could the variables provide on the whole process. In the same quadrant, the variables which were adjacent to each other were positively correlated; meanwhile, the variables in different quadrant were negatively correlated. As shown in Fig. 1, the total sugar yield, glucan yield, xylan yield, the hydrolysis efficiency of glucan, the hydrolysis efficiency of xylan, the total amount of alkaline, the removal of lignin were close to each other in the same quadrant, which showed that there was a positive correlation between these variables. A clear negative correlation example was the lignin content and the delignification rate. The two variables were far away from each other in different quadrant. Obviously, when the D-lignin value was higher, the lignin content in the pretreated corn stover was lower. A large amount of lignin was removed during the pretreatment, as a result, after alkaline pretreatment; the amount of lignin ingredient of corn stover samples was relatively small.

Seen from the P1 perspective, during the alkaline pretreatment process, the total amount of alkaline had a very important influence on the delignification results. The actual experimental results could also show that the greater the total alkaline amount, the greater the removal rate of lignin would get. The lignin was removed thoroughly and the value of lignin content of the pretreated samples would be low.

Seen from the P2 perspective, during the alkaline pretreatment process, the relationship of pretreatment temperature and the holding time was positive correlation. To some extent, temperature and time could make up for each other. The conditions of shorter holding time at relatively higher temperature and longer holding time at relatively lower temperature could get similar pretreatment effect. Pretreatment temperature, holding time and alkaline charge had positively correlated with lignin removal rate, but had negative correlation with lignin removal rate, recovery rate of glucan and recovery rate of xylan. The main reason was that higher temperature, longer holding time and higher total alkaline charge could cause the degradation of carbohydrates, breaking and removing of the macromolecular lignin chains. Thus, these pretreatment conditions could result in the lower value of total sugar yield, glucan yield and xylan yield.

In the scatter plot figure of P3 and P4 (Fig. 1c), about 19% information could be explained by the third components and the fourth components. Seen from the P3 and P4 perspective, holding time and temperature had negative correlation. The yield of Xylan had positive correlation with xylan content, but had negatively correlated with temperature. The total sugar yield, glucan yield, xylan yield, glucan content and the recovery of glucan were close to each other in the same quadrant, which showed that there was a positive correlation between these variables.

The cumulated R2 and Q2 values for each variable of PCA could be seen in Fig. 1d. Most of variables have high green bars, R2, and high blue bars, Q2. Here, R2 indicates how well the variation of a variable is explained, while Q2 indicates how well a variable can be predicted. Thus, variables were well modeled in PCA. Based on
Fig. 1. PCA analysis.
(a: Score scatter plot; b: Loading scatter plot for p1 and p2; c: Loading scatter plot for p3 and p4; d: overview graphics (for each variable) of the values of cumulated R2 and Q2.)

Note: The overall graph of PCA analysis based on the entire dataset. a: the scores t1, t2, t3 and t4, were summarizing the variables. The scores were orthogonal, and they were independent of each other, completely. The first component was showed by t1. t1 could showed the largest variation data information of X followed by t2, t3, t4. b: p1 (for first component) and p2 (for second component) showed the weights combining the original variables. c: p3 (for third component) and p4 (for fourth component) showed the weights combining the original variables. d: The fraction of the variation explained by this PCA model was presented by R2. This model could predict some fraction of variations and it could be valued by Q2.
the PCA analysis, 4 PLS models were built (discussed in the following) and the coefficient values have been calculated to illustrate the relationship between different parameters. The coefficient values of different PLS model could be seen in Supporting Information Fig. 1.

3.2. PLS model 1 analysis for the whole process

As shown in Fig. 2, 13 variables of the pretreatment parameters were picked from the 18 variables for the PLS model 1 analysis, which including 5 independent variables (X) and 8 dependent variable (Y). The Alkaline, S/L ratio, T, H time and PFI was set as X, and the chemical composition and pretreatment effect parameters (Glucan, Xylan, Extractive, Lignin, Solid Yield, D-lignin, R-glucan, and R-xylan) were set as Y. The PLS model 1 showed that for the dependent variable X parameters, R2X(cum) was 0.800, the R2Y (cum) was 0.655 for the independent variable Y. The square of Q was 0.517. In other words, in the PLS model 1, the two main component could explain the 65.5% information of the pretreatment process based on the 13 variables. The predictive ability of the PLS model 1 was 52%. The results showed that the PLS model 1 based on the 13 variables had a relatively good assessing capability.

As can be seen in the figure, the PFI variables were at the origin of the coordinate graphs, indicating that there was no relationship between the PFI variable and other variables in this PLS model. The main reason was that PFI was a kind of mechanical refining treatment after the alkaline pretreatment of the corn stover, and there was no chemical reaction or chemical charge during the PFI treatment process. Generally speaking, after PFI mechanical refining treatment, the significant difference of corn stover samples reflected on the size, shape, and surface appearance. In the chemical composition, such as lignin content, extractive content, glucan content, xylan content, and in the pretreatment effect, such as recovery rate of glucan, xylan, and removal rate of lignin, the corn stover samples with and without PFI treatment provided the same values.

As seen in the figure, the positive correlation and negative correlation between the variables showed in the both vertical and horizontal coordinate axis were the same with the correlations in the PCA model analysis. Holding time and the pretreatment temperature were closely related to each other. Alkaline charge and removal rate of lignin showed positive correlations. Solid yield, recovery rate of glucan, recovery rate of xylan, extractive negatively correlated with alkaline charge. In view of the distance between the variable and the origin point, the impact of I/S ratio on the whole pretreatment process was smaller than the other variables. As could be seen in the VIP graph of X (the importance of independent variables), during the pretreatment process, the greatest impact was from the variable of alkaline charge, which provided more important influence than any other process variables.

With the increase of the total alkaline charge, the removal rate of lignin improved significantly. For example, when the liquid to solid ratio was 6:1, temperature was 130 °C, the removal rate of lignin was 21.1%, 33.72%, 51.7%, 73.9% and 77.54% respectively, and the solid recovery rate was 68.16%, 59.82%, 59.87%, 50.61% and 55.52%, respectively, according to the total alkaline charge (5, 7, 9, 11, 13%). Obviously, with the increase of the total alkaline charge, the solid recovery rate reduced. The main reason was that the alkaline could promote the degradation of carbohydrate while removed the lignin. So there was a certain loss of glucan and xylan which led to the reduction of solid recovery rate. As can be seen in the experimental results, when the alkaline charge was 5%, the recovery rate of glucan was 88.45%, and when the alkaline charge increased to 11%, the recovery rate of glucan was decreased to 83.63%. Meanwhile, the recovery rate of xylan was decreased from 80.70% to 65.43%.

In addition of total alkaline charge, the second, third, fourth importance indexes (VIP) were temperature of pretreatment, pretreatment time (holding time at the highest temperature), and ratio of liquid to solid, respectively. As discussed above, PFI had no contribution to variables in this model, and its VIP was zero.

3.3. The PLS model 2 analysis for the whole process

As shown in Fig. 3, in the PLS model 2 analysis, the pretreatment condition parameters were set as X, the chemical component, pretreatment effect, and the hydrolysis effect were set as Y. There were 18 variables in the PLS model 2, including 5 variables for X, and 13 variables for Y. Alkaline, S/L ratio, T, H time, PFI were set as X, Glucan, Xylan, Extractive, Lignin, Solid Yield, D-lignin, R-glucan, R-xylan, E-glucan, E-xylan, Y-glucan, Y-xylan, Y-total sugar were set as Y. In this PLS model 2, the R2X(cum) was 0.401, R2Y(cum) was 0.536, Q2 was 0.436.

Similar with PLS model 1, PLS model 2 showed the relatively similar results. In the diagram layout, the holding time and pretreatment temperature were close to each other, and away from the other variables point in other quadrants. The main reason was that the holding time and pretreatment temperature had negative correlation with the pretreated corn stover ingredients, pretreatment efficiency and enzymatic efficiency. To some extent, the longer holding time and higher pretreatment temperature could bring negative impact on the final enzymatic effect. E-glucan, E-xylan, Y-glucan, Y-xylan, Y-total sugar. Xylan, Glucan had a strong positive correlation between each other, so they were very close to each other in the PLS model 2 graph. When the glucan and xylan content were higher in the pretreated corn stover samples, the enzymatic saccharification process could get better effect, and the hydrolysis efficiency of glucan and xylan could be improved, which also improved the final total sugar yield.

As can be seen in the PLS model 2 figure, the variables such as Solid yield, R-xylan, R-glucan, Lignin, Extractive were close to each other, which reflected the close relationship between these variables. This result could be explained by the relationship between the component of pretreated corn stover samples and the solid recovery rate. The solid recovery rate was the ratio of weight of pretreated samples to the weight of raw corn stover samples. Obviously, the weight of pretreated samples was decided by the content of each ingredient of the corn stover after alkaline pretreatment. As a result, when the content of lignin, extractive, glucan, xylan was relatively high, the recovery rate of glucan and xylan could get high value, and the total weight of pretreated samples would be high, which could led to the higher solid recovery rate of the whole process.

Compared with PLS model 1, one of the main differences was that PFI was not in the origin of PLS model 2. In this model, the change of PFI variable could influence the other variables. As can be seen in the two main components of the graph, PFI had good positive correlation with E-glucan, E-xylan, Y-glucan, Y-xylan, Y-total sugar. In the VIP analysis graph, the PFI variable could also have contribution to the change of other variables results. For example, when the S/L ratio was 6, temperature was 130 °C, the holding time was 30 min, the total alkaline charge was 11%, the Y-glucan, Y-xylan, Y-total sugar without PFI treatment were 79.49%, 52.76%, 69.65% respectively, however, the Y-glucan, Y-xylan, Y-total sugar with PFI treatment were 83.6%, 58.6%, 74.4% respectively. Obviously, after PFI pretreatment, the Y-glucan, Y-xylan and Y-total sugar were increased, and increased by 4.11%, 5.84% and 4.75% respectively (all original pretreatment and enzymatic hydrolysis data could be seen in Supporting Information Table 2). This showed that the PFI treatment in the whole pretreatment of corn stover.
Note: In this model, Alkaline, S/L ratio, T, H time and PFI was set as X, and the chemical composition and pretreatment effect parameters (Glucan, Xylan, Extractive, Lignin, Solid Yield, D-lignin, R-glucan, and R-xylan) were set as Y. a: the scores $t_1$ and $t_2$, were summarizing the variables. The scores are orthogonal, completely independent of each other; the score $t_1$ (first component) explains the largest variation of the X space, followed by $t_2$. b: $w^c[1]$ (for first component) and $w^c[2]$ (for second component) showed the weights combining the original variables. c: the VIP (Variable Importance for the Projection) plot summarized the importance of the variables both to explain X and to correlate to Y. VIP plots are sorted by order of importance.
Fig. 3. PLS model 2 analysis (a: score scatter plot, b: loading scatter plot, c: Variable Importance for the Projection).
(Alkaline, S/L ratio, T, H time, PFI were set as X, Glucan, Xylan, Extractive, Lignin, Solid Yield, D-lignin, R-glucan, R-xylan, E-glucan, E-xylan, Y-glucan, Y-xylan, Y-total sugar were set as Y.)

Note: a: the scores $t_1$ and $t_2$ were summarizing the variables. The scores are orthogonal, completely independent of each other, the score $t_1$ (first component) explains the largest variation of the X space, followed by $t_2$. b: $w^c[1]$ (for first component) and $w^c[2]$ (for second component) showed the weights combining the original variables. c: the VIP (Variable Importance for the Projection) plot summarized the importance of the variables both to explain X and to correlate to Y. VIP plots are sorted by order of importance.
Fig. 4. PLS model 3 analysis (a: score scatter plot, b: loading scatter plot, c: Variable Importance for the Projection).
(The pretreatment parameters, the chemical compositions of pretreated corn stover samples and the pretreatment effect were set as X variables, and the enzymatic hydrolysis was set as Y variables. Alkaline, S/L ratio, T, H time, PFI, Glucan, Xylan, Extractive, Lignin, Solid Yield, D-lignin, R-glucan, R-xylan were set as X variables, meanwhile, E-glucan, E-xylan, Y-glucan, Y-xylan, Y-total sugar were set as Y variables.)

Note: a: the scores t1 and t2, were summarizing the variables. The scores are orthogonal, completely independent of each other, the score t1 (first component) explains the largest variation of the X space, followed by t2. b: \( w^c[1] \) (for first component) and \( w^c[2] \) (for second component) showed the weights combining the original variables. c: the VIP (Variable Importance for the Projection) plot summarized the importance of the variables both to explain X and to correlate to Y. VIP plots are sorted by order of importance.
Fig. 5. PLS model 4 analysis (a: score scatter plot, b: loading scatter plot, c: Variable Importance for the Projection).
(Glucan, xylan, Extractive, Lignin, Solid Yield, D-lignin, R-glucan, R-xylan were set as X variables, E-glucan, E-xylan, Y-glucan, Y-xylan, Y-total sugar were set as Y variables).
Note: a: the scores t1 and t2, were summarizing the variables. The scores are orthogonal, completely independent of each other. the score t1 (first component) explains the largest variation of the X space, followed by t2. b: \( w[c][1] \) (for first component) and \( w[c][2] \) (for second component) showed the weights combining the original variables. c: the VIP (Variable Importance for the Projection) plot summarized the importance of the variables both to explain X and to correlate to Y. VIP plots are sorted by order of importance.
Fig. 6. Influence of independent variables on the variables.
(a: E-glucan. b: E-xylan. c: Y-glucan. d: Y-xylan. e: Y-total sugar)
(The influence degree of all the independent variables X on the each dependent variables Y.)
could improve the hydrolysis efficiency, improve the yield of glucan and xylan, which could led to increase the final total sugar yield.

3.4. The PLS model 3 analysis for the whole pretreatment process

As shown in Fig. 4, in the PLS model 3, there were 18 variables, which including 13 X variables and 5 Y variables. The pretreatment parameters, the chemical compositions of pretreated corn stover samples and the pretreatment effect were set as X variables, and the enzymatic hydrolysis was set as Y variables. Alkaline, S/L ratio, T, H time, PFI, Glucan, Xylan, Extractive, Lignin, Solid Yield, D-lignin, R-glucan, R-xylan were set as X variables, meanwhile, E-glucan, E-xylan, Y-glucan, Y-xylan, Y-total sugar were set as Y variables. In this PLS model, after two principal components analysis, the $R^2_X$(cum) was 0.544, the $R^2_Y$(cum) was 0.643, $Q^2$(cum) was 0.585. This PLS model could explain the 64% information of the process, and had 59% predictive ability.

As can be seen in the model graph, the variables which could provide great impact on the enzymatic efficiency were as follows: the total alkaline charge, removal rate of lignin, lignin content, glucan content, xylan content, solid recovery, pretreatment temperature, extractive, the ratio of liquid to solid. Wherein, the total alkaline charge, removal rate of lignin and lignin content had the similar importance degree, which showed that the lignin removal had very important influence on the subsequent enzymatic saccharification after alkaline pretreatment. The lignin removing was mainly got by the reaction of lignin and alkaline chemicals during the pretreatment.

For example, when the removal rate of lignin was 21.1%, the E-glucan, E-xylan, Y-glucan, Y-xylan and Y-total sugar were 58.9%, 42.4%, 59.1%, 41.2%, 52.51%, respectively. When the removal rate of lignin increased to 51.7%, the E-glucan, E-xylan, Y-glucan, Y-xylan and Y-total sugar were increased to 72.4%, 52.9%, 70.0%, 55.9%, and 64.8% respectively. When the removal rate of lignin was further increased to 77.54%, the E-glucan, E-xylan, Y-glucan, Y-xylan and Y-total sugar were increased to 78.05%, 51.90%, 77.38%, 53.19%, and 68.48% respectively. However, it should be noted that the increase of the removal rate of lignin could not necessarily correspond to the increase of the enzymatic efficiency. That was because the enzymatic hydrolysis was a complex process, which was subject to many factors. When the removal rate of lignin was too high, there could be a great loss of carbohydrates which could decrease the total sugar yield in the enzymatic hydrolysis process.

3.5. The PLS model 4 analysis for the whole process

In the PLS model 4, there 13 variables, which including 8 X variables and 5 Y variables for this analysis. The chemical composition of pretreated corn stover, pretreatment effect were set as X variables (Glucan, xylan, Extractive, Lignin, Solid Yield, D-lignin, R-glucan, R-xylan), and the enzymatic hydrolysis effect variables were set as Y variables (E-glucan, E-xylan, Y-glucan, Y-xylan, Y-total sugar). After the two principle components analysis of the PLS model, the $R^2_X$(cum) was 0.739, the $R^2_Y$(cum) was 0.564, $Q^2$(cum) was 0.514. This PLS model 4 could explain the 56.4% of information of the experimental data, and its predictive ability was 51.4%.

As can be seen in the VIP graph of Fig. 5, the X independent variables which had great influence degree on the overall Y dependent variables were as follows: removal rate of lignin, lignin content, glucan content, xylan content, solid recovery rate, glucan recovery rate, xylan recovery rate and extractive content. The results showed that the main four components of corn stover (glucan, xylan, lignin and extractive) had great influence on the final total sugar yield, especially the lignin had the greatest influence and the extractive had the lowest influence.

The overall VIP analysis results showed that in order to get high enzymatic effect, the lignin of raw corn stover samples should be removed. Lignin removing could help to take over the main obstacle of enzymatic hydrolysis after pretreatment. At the same time, during the alkaline pretreatment, the carbohydrates (glucan and xylan) should be reserved as much as possible, which could supply the fermentable sugar during the subsequent enzymatic hydrolysis process.

In this PLS model, the influence degree of all the independent variables X on the each dependent variables Y could be seen in the analysis graph. For example, the hydrolysis efficiency of glucan and xylan mainly had strong positive correlation with the glucan content, xylan content, and removal rate of lignin. Meanwhile, they had negative correlation with solid recovery rate and lignin content, and had no obvious relations between recovery rate of glucan and xylan. The reasons were as follows: extractive had little effect on the process mainly because most of the extractive had been removed after alkaline pretreatment.

As shown in Fig. 6, based on the PCA and PLS model analysis above, in the sodium hydroxide based alkaline pretreatment of corn stover, in order to improve the subsequent enzymatic saccharification effect after pretreatment, the alkaline charge could be increased, the maximum pretreatment temperature should be decreased, the holding time of the maximum pretreatment temperature should not be longer (more than 30 min in this paper). The analysis results also showed that PFI treatment could improve the final total sugar yield in the whole process.

4. Conclusions

For the pretreatment of corn stover, 18 variables were analyzed by PCA and PLS analysis. The PCA analysis results provided the relationship between the 18 variables. By setting different independent variables and dependent variables, the PLS model analysis provided the influence degree of the dependent variables on the independent variables. The final analysis results showed that total alkaline charge, and removal rate of lignin had greatest impact on the pretreatment and enzymatic hydrolysis. The ratio of liquid to solid and the extractive content had a little impact on the final experimental results. The PFI treatment had obviously positive correlation with the final enzymatic hydrolysis effect.

After the PCA and PLS analysis, the relationship between the pretreatment condition parameters, composition of materials, pretreatment effect and enzymatic effect could be provided, and the correlations in-between could play a significant role in guiding the whole process of pretreatment of corn stover. The PCA and PLS analysis could help to optimize conditions to improve the final enzymatic saccharification efficient after alkaline pretreatment of lignocellulosic biomass.

Acknowledgements

This work was also partially supported by the National Natural Science Foundation of China (No. 31470609), the Qingdao Institute of Bioenergy and Bioprocess Technology Director Innovation Foundation for Young Scientists (No. Y37205110C), the Qingdao University of Science and Technology Foundation for Young Teachers (No. 010022706), and the Open Foundation of CAS Key Laboratory of Bio-based Materials (QIEBT-CAS).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.resconrec.2016.12.007.
Reference:

Ågbor, V.B., Ciccek, N., Sparling, R., Berlin, A., Levin, D.B., 2011. Biomass pretreatment: fundamentals toward application. Biotechnol. Adv. 29 (6), 675–685.

Ayamga, E.A., Kemausuor, F., Addo, A., 2015. Technical analysis of crop residue biomass energy in an agricultural region of Ghana. Resour. Conserv. Recycl. 96, 51–60.

Bichraoui-Draper, N., Xu, M., Millerb, S.A., Guillaumea, B., 2015. Agent-based life cycle assessment for switchgrass-based bioenergy systems. Resour. Conserv. Recycl. 103, 171–178.

Ding, S.Y., Liu, Y.S., Zeng, Y., Himmel, M.E., Baker, J.O., Bayer, E.A., 2012. How does plant cell wall nanoscale architecture correlate with enzymatic digestibility? Science 338, 1055–1059.

Godoy, J.L., Vega, J.R., Marchetti, J.L., 2014. Relationships between PCA and PLS-regression [J]. Chemom. Intell. Lab. Syst. 130, 182–191.

Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W., Foust, T.D., 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315, 804–807.

Khoo, H.H., Wong, L.L., Tan, J., Isoni, V., Sharratt, P., 2015. Synthesis of 2-methyl tetrahydrofuran from various lignocellulosic feedstocks: sustainability assessment via LCA. Resour. Conserv. Recycl. 95, 147–182.

Li, B., Bandekar, R., Zha, Q., Alsagag, A., Ni, Y., 2011. Fiber quality analysis: OpTest fiber quality analyzer versus L&W fiber tester. Ind. Eng. Chem. Res. 50 (22), 12572–12578.

Li, B., Liu, H., Xu, H., Pang, B., Mou, H., Wang, H., Mu, X., 2014. Characterization of the detailed relationships of the key variables in the process of the alkaline sulfate pretreatment of corn stover by multivariate analysis. BioResources 9 (2), 2772–2771.

Lin, W.S., Lee, M.S., Huang, Y.C., Den, Walter, 2015. Identifying water recycling strategy using multivariate statistical analysis for high-tech industries in Taiwan. Resour. Conserv. Recycl. 94, 35–42.

Liu, C., Van der Heide, E., Wang, H., Li, B., Yu, G., Mu, X., 2013. Alkaline twin-screw extrusion pretreatment for fermentable sugar production. Biotechnol. Biofuels 6, 97.

Ljungqvist, C.H., Lyng, R., Thuvander, F., 2005. Evaluation of PFI beating on the strain to failure of spruce fibres using single fibre fragmentation. Nord Pulp Pap. Res. J. 20 (4), 370–377.

Luo, X.L., Zhu, J.Y., 2011. Effects of drying-induced fiber hornification on enzymatic saccharification of lignocelluloses. Enzyme Microb. Technol. 48, 92–99.

Luo, X.L., Zhu, J.Y., Gleiser, N., Zhan, H.Y., 2011. Effects of wet-pressing-induced fiber hornification on enzymatic saccharification of lignocelluloses. Cellulose 18, 1055–1062.

Miura, T., Lee, S.-H., Inoue, S., Endo, T., 2012. Improvement of enzymatic saccharification of sugarcane bagasse by dilute-alkali-catalyzed hydrothermal treatment and subsequent disk milling. Bioresour. Technol. 105, 95–99.

Oh, S.Y., Dong, I.Y., Shin, Y., Hwan, C.K., Hak, Y.K., Yong, S.C., Won, H.P., Ji, H.Y., 2005. Crystalline structure analysis of cellulose treated with sodium hydroxide and carbon dioxide by means of X-ray diffraction and FTIR spectroscopy. Carbohydr. Res. 340, 2376–2391.

Paranhos, R.S., Petter, C.O., 2013. Multivariate data analysis applied in Hot-Mix asphalt plants. Resour. Conserv. Recycl. 73, 1–10.

Segal, L., Creely, J., Martin, A., Conrad, C., 1959. An empirical method for estimating the degree of crystallinity of native cellulose using the X-ray diffractometer. Text. Res. J. 29 (10), 786–794.

Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., Templeton, D., Crocker, D., 2008. Determination of Structure Carbohydrates and Lignin in Biomass. Laboratory Analytical Procedure (LAP). NREL/TP-510-42618. National Renewable Energy Laboratory, Golden, Colorado, USA.

Sun, Y., Cheng, J.Y., 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review [J]. Bioresearch. Technol. 83 (1), 1–11.

Van Dyk, J.S., Pletschke, B.L., 2012. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-factors affecting enzymes, conversion and synergy[J]. Biotechnol. Adv. 30, 1458–1480.

Von Schenck, A., Berglin, N., Uusitalo, J., 2013. Ethanol from Nordic wood raw material by simplified alkaline soda cooking pre-treatment [J]. Appl. Energy 102, 229–240.

Yang, L.F., Cao, J., Mao, J.Y., et al., 2013. Sodium carbonate-sodium sulphate pretreatment for improving the enzymatic hydrolysis of rice straw [J]. Ind. Crop Prod. 43, 711–717.

Yu, G., Li, B., Liu, C., Zhang, Y., Wang, H., Mu, X., 2013. Fractionation of the main components of corn stover by formic acid and enzymatic saccharification of solid residue. Ind. Crop Prod. 50, 750–757.

Zhu, J.Y., Pan, X.J., Wang, G.S., Gleiser, R., 2009. Sulfite pretreatment (SPORL) for robust enzymatic saccharification of spruce and red pine. Bioresearch. Technol. 100 (8), 2411–2418.