Rapid Screening of Microfibrillated Cellulose Tructure Through FTIR and Principal Component Analsysis

To cite this article: N F N M Nor *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1192** 012009

View the article online for updates and enhancements.

Rapid Screening of Microfibrillated Cellulose Tructure Through FTIR and Principal Component Analsysis

N F N M Nor[1], N F M Azmin[1*], A L Asnawi[2*], D N Jimat[1], S Abdullah[3]

[1]Department of Biotechnology Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia
[2]Department of Electrical Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia
[3]Faculty of Chemical and Process Engineering Technology, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Kuantan, Pahang, Malaysia

*Corresponding author's email: norfadhillah@iium.edu.my; aniliza@iium.edu.my

Abstract. Analysis of FTIR spectra combined with multivariate statistical analysis technique specifically Principal Component Analysis was used for rapid screening of microfibrillated cellulose (MFC) structure. The current methods used to extract the MFC are by using the chemical and physical approaches. To date, most researchers focused on bench (lab) scale experiment to identify the structure of MFC. Lack of mathematical models focusing on this goal has motivated this project. Principal component analysis is applied to identify the chemical composition of the MFC. The dataset comprises FTIR spectra of 12 samples that comes from MFC with different particles sizes, 200 µm, 250 µm and 800 µm. The result shows that the wavelength region which represents the MFC structure is in the range of 2950 cm^{-1} to 2978 cm^{-1} for particle size of 200 micrometer since it has larger surface area for penetration of fungal into the biomass due to lower diffusion of air, water and metabolite intermediates of which cellulose can be easily hydrolyzed due to increase in pore size of substance through greater removal of hemicellulose and lignin. The overall result indicates that the combination of FTIR analysis and PCA is a useful technique for rapid screening of MFC structure.

1. Introduction

Microfibrillated cellulose (MFC) can be referred as a larger unit of approximately more than thirty individual cellulose molecules brought together by biomass [1]. It is obtained through fibrillation of lignocellulose materials [2]. The current methods for screening of MFC structure are slow, laborious, and use harsh reagents which are impractical to use in screening process. This necessitates the search for alternative methods and one of the under exploration is an approached called chemometrics. Chemometric and Intelligent Laboratory Systems, the leading journal in the field of chemometric defines chemometric as the chemical discipline that uses mathematical, statistical, and other methods employing formal logic design or select optimal measurement procedures experiments, and to provide maximum relevant chemical information by analysing chemical data [3].

FTIR spectroscopy was selected in this study due to its simplicity, rapid and non-destructive measurements of chemical and physical compositions, high sensitivity and selectivity as well as real

time detection on the effect of fungal pre-treatment and ultrasonication for different particle sizes of the MFC structure. As shown by Amir [4] in their study to identify wheat varieties, the FTIR analysis enables them to quickly visualize the chemical composition and estimate the quality parameters of different types of wheat.

However, additional data processing technique is required due to the complexity in unmasking the spectra. Application of the multivariate statistical technique for the analysis of the FTIR data was necessary in order to perform feature extraction process, proper evaluation, and identification of the obtained spectra. Process that involves analysing the biomass structure and monitoring changes that occur from physical, chemical or biologicaltreatments currently required the non-destructive analytical tool which generated hundreds of spectra within short period of time [5]. Basically, pre-treatment process aimed to remove lignin and hemicellulose, reduce crystallinity of cellulose as well as increase the porosityof the lignocellulosic materials [6,7]. This study aimed to identify the effect of particle sizes on the screening process since large particle size will limit the penetration of fungi into biomass due to lower diffusion of air, water and metabolite intermediates.

Principal Components Analysis (PCA) involves the establishment of the grouping structures of various lignocellulosic biomasses by transformation of the correlated variables in the multivariate data into the uncorrelated variables which is principal components (PCs).The main goals in PCA can be illustrated as the process of extracting the important information from data tables in order to reduce the size of data by providing the important information that would simplified the set of data based on the analysis of the observations and variables structure. In PCA, important data is express as principal components in which it represents a set of orthogonal variables. Thus, the combination of FTIR and multivariate statistical analysis are well suited for identification and differentiation purposes even in very large data sets.

2. Materials and Methods

2.1. Raw Data
FTIR analysis was carried out from run 1 until run 12 with different particles size and inoculum concentrations. All samples were triplicate in SSF making the total of 12 samples with two independent variables particularly granular raw particle size (RPS) at 200, 525, and 850 μm and initial inoculum size (IS) with the compositions of 2.0, 3.5, and 5.0% (w/v) of CPH.

2.2. Data Pre-processing
The result obtained from FTIR analysis was pre-processed by organizing the spectral data to obtain the appropriate data matrix subjected to the PCA algorithm. FTIR dataset consists of 12 samples. The dataset was transposed and as a result, the rows represent the wavenumbers whilst the columns represent the samples of the MFC. Direct data matrix is not commonly practice in characterization of MFC structure since the components was analysed according to the variation of the absorbance frequency in each band while in transposed matrix, factors are computed based on the intensity of absorbance relative to the whole spectrum. Hence the identification of MFC structure was performed by applying PCA on the transposed matrix.

2.3. Principal Component Analysis
PCA was executed using MATLAB software using the following steps.

Step 1: Normalization of the dataset
Each row in the dataset was normalised using the zscore function in MATLAB's library.

Step 2: Obtaining the eigenvalues and eigenvectors
Eigenvalues and eigenvectors were obtained from the covariance matrix. Variabilities in the data were explained by the magnitude of eigenvalues. Scree plot was used to explain the data variability. The x-

axis in scree plot represents the index of the number of eigenvalues whereas y-axis denoted as magnitude of the eigenvalues. Thus, values in vector Y must not include any non-negative function and NaNs.

Step 3: Generating plots from the principal components analysis
A scree plot, score plots and loading plot were generated subsequent to the PCA to enables extraction of information.

3. Results

3.1. Scree Plot
The scree plot enables the determination of number of principal components to be kept or discarded because it indicates the relative importance of each principal component. The number of principal components retained should represents at least 70 to 80 percent of the variance in the data. In this study, only the first two principal components. PC1 and PC2, were selected while the rest of the principle components can be discarded as shown in Figure 1. PC1 explained 92% variability of the data and PC2 explained 2% of data variability. Total percentage of variance explained by PC1 and PC2 is 94% in which within the acceptable range for a validated result.

Figure 1. Scree plot of the data.
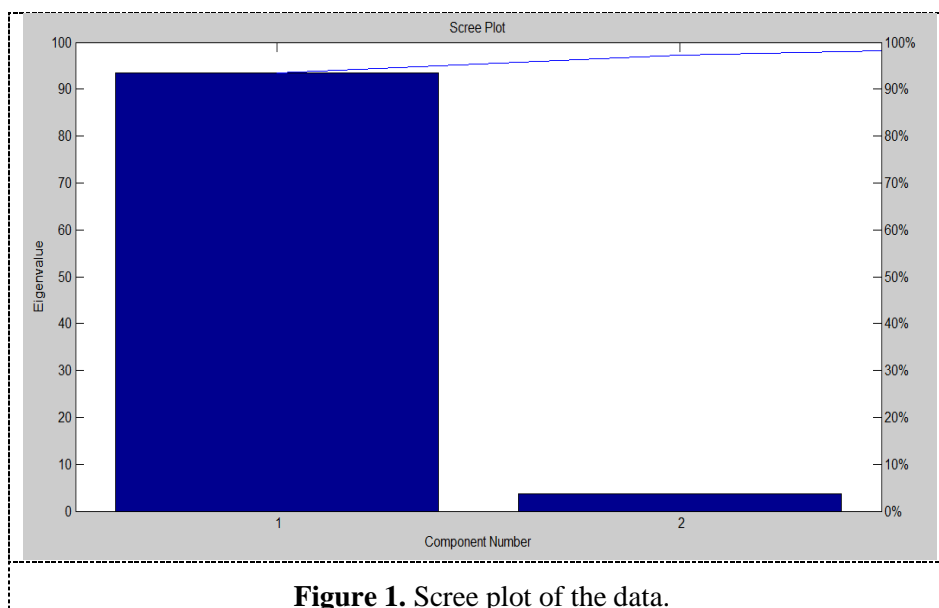
3.2. Score Plot
Score plot indicates the relationship between the samples by providing the summaries of the observations. For PC1, the samples can be grouped into two groups which are the positive region and negative region. Sample 1,3,4,5 and 9 can be group into one group which is positive region in PC1 as they are lie on the positive region of PC1 while sample 2, 6, 7, 8, 10, 11 and 12 lie on the negative side of PC1. So, they can be grouped into one group. Sample 1, 3 and 4 have similarity in their particle size which is 200 μm. Sample 6, 7 and 8 represents the sample with particles size of 525 μm. Particle size of 850 μm can be represented by the sample 10, 11 and 12. The control samples denoted as sample 1, 5 and 9. Based on the observation and inference, a conclusion can be made for PC1 which it is distinguishing the particles size of 200 μm with particles sizes of 525 μm and 850 μm with the exception

of sample 2 which is lie on negative side of PC1 due to less amount of fungus penetrate into the biomass in which it provide inefficient degradation of hemicellulose and lignin.

Figure 2. Score plot of PC1 versus PC2.
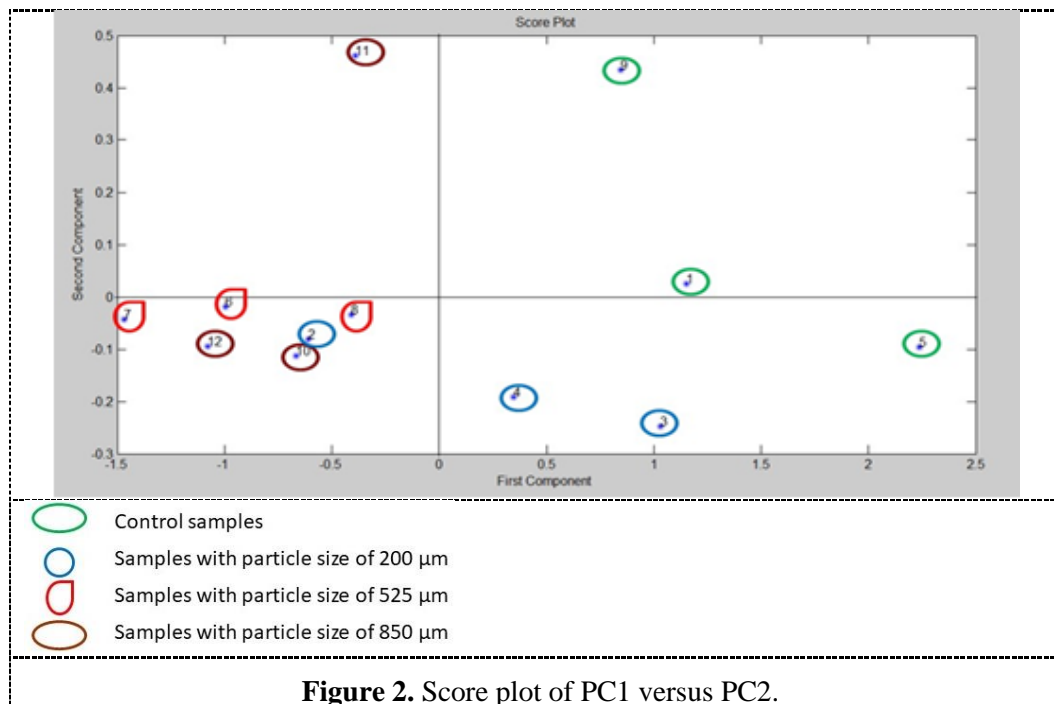
3.3. Loading Plot

This loading plot provides relationship between all variables. In loading plot, each variable is represented as a vector while direction and length of the vector indicate how variable contributes to the principal components in the plot. Figure 3 represents the bi-plot of which it consists of both score and loading of a selected component.

Based on this loading plot, it can be observed that MFC structure was presence within the wavelength ranging from 2950 cm-1 to 2978 cm^{-1}. It can be clearly seen that the particle size of 200 μm indicate the presence of MFC structure in which it is important variable to PC1 and the sample of run 3 and 4 are closest to this region of MFC structure since these samples have highest correlation with the wavelength representing the region of MFC structure. FTIR data currently focused on the optimization of MFC from cocoa pod husk (CPH) substrate in two distinct processes. Substrate was subjected to fungus-pre-treatment at the first stage through a solid-state fermentation by *Aspergillus nig*er to promote enzymatic splitting of cellulose. The fungus pre-treated substrate was then mechanically processed by the application of high-intensity ultrasonication. The optimum process of the fungus pre-treatment can be obtained from samples 3 and 4 of 200 μm since in these samples; fungus facilitated the splitting and reducing size of individual fiber into the microfibrils size. It can be concludes that particle size of 200 μm have larger surface area for penetration of fungus into the biomass due to lower diffusion of air, water and metabolite intermediates in which cellulose can be easily hydrolysed due to increase in pore size of substrate through greatly removal of hemicellulose and lignin. Thus, fungal pre-treatment provides a significant effect towards the MFC particle size produced by providing the required enzyme

of cellulase to assist the ultrasonic process to split and reduce the size of individual fiber into the microfibrils size.
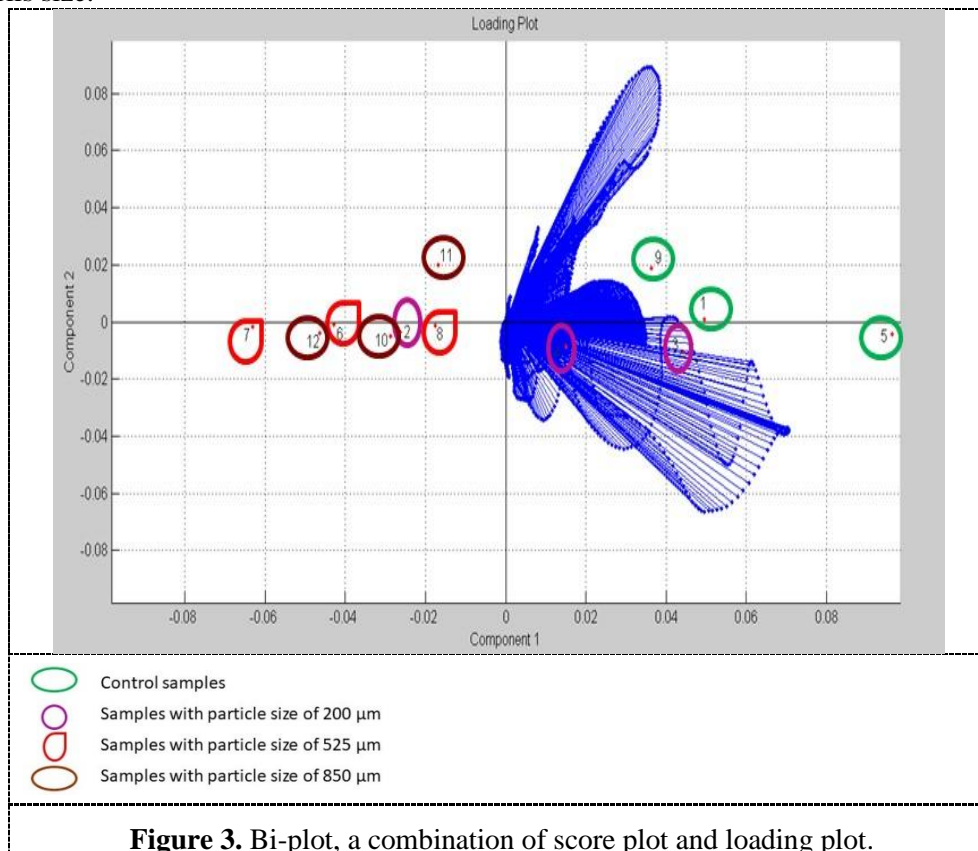

Figure 3. Bi-plot, a combination of score plot and loading plot.

4. Discussion

Based on the scree plot obtained, only the first two principal components, PC1 and PC2 were selected while the rest of the principal components were discarded. Remaining subsequent principal components explained percentages less than 5. PC1 explained 92% of the data variance and whereas PC2 explained 2%. Total percentage of variance explained by PC1 and PC2 is 94% which within the acceptable range for a validated result. The reliability of the analysis depends on the variance explained and the number of principal component (PCs) used.

Rapid screening of the MFC in cocoa pod husk biomass is based on score plot and divided according to the sample types in order to provide further understanding on the relations between them. The division is based on the number of runs with different particles size through combinations of fungal pre-treatment with the high-intensity ultrasonification. Samples were divided into control runs and actual runs. A control run indicates the sample undergoes ultrasonification without fungal pre-treatment. Basically, there are 3 control runs for 3 different particle sizes. There are 9 actual runs in which it undergoes fungal pre-treatment and ultrasonification. These runs are categorized equally according to 3 different particle sizes. Ultrasonification reduces the size of biomass by altering structure of hemicellulose and lignin. Efficiency of the ultrasonic pre-treatment depends on the size of particles. Samples obtained through combination of fungal pre-treatment and ultrasonification will remove the major portion of the hemicellulose and lignin.

For PC1, the samples can be grouped into two groups which are the positive region and negative region. Sample 1,3,4,5 and 9 can be grouped into one group which is positive region in PC1 as they lie on the positive region of PC1 while sample 2, 6, 7, 8, 10, 11 and 12 lie on the negative side of PC1. So, they can be grouped into one group. Sample 1, 3 and 4 have similarity in their particle size which is 200

μm. A conclusion can be made for PC1 which it is distinguishing the particles size of 200 μm with particles sizes of 525 μm and 800 μm with the exception of sample 2 which is lie on negative side of PC1 due to less amount of fungal penetrate into the biomass in which it provides inefficient degradation of hemicellulose and lignin.

Similar to PC1, PC2 also can be grouped into positive region and the negative region. Only sample 1,9 and 11 are in the positive region of PC2. Sample 9 and 11are located closed to each other because they are underlying within similar particles sizes which are 850 μm. Sample 1 is located far away from sample 9 and 11 because they are different types of particles size. Based on the observation that have been made, PC2 differentiates the particles size of 850 μm with the rest of the other particles' sizes with the exception of sample 10 and 12 in which it lies between negative region of PC2. This might due to the increasing in particle size result in smaller surface area to volume ratio in which cellulose structure is more accessible for degradation by fungal.

Based on the loading plot, it can be observed that MFC structure was presence within the wavelength ranging from 2950 cm^{-1} to 2978 cm^{-1}. It can be clearly seen that the particlesize of 200 μm indicate the presence of MFC structure in which it is important variable to PC1 and the sample of run 3 and 4 are closest to this region of MFC structure since these samples have highest correlation with the wavelength representing the region of MFC structure. It can be concluded that particle size of 200 μm have larger surface area for penetration of fungal into the biomass due to lower diffusion of air, water and metabolite intermediates in which cellulose can be easily hydrolyzed due to increase in pore size of substrate through greatly removal of hemicellulose and lignin.

5. Conclusion

FTIR analysis is an infeasible approach in screening of MFC since large number of spectral involves and hardly distinguishable by visual inspection. Integration between FTIR analyses with PCA analysis will establish the underlying groups' structure of lignocellulose biomass in order to improve the screening process. Algorithm will record the peak present in each spectrumincluding the presence of their functional groups as well as corresponding peak area. Thus,PCA analysis provides a relationship between the effects of particle sizes on the screening process as the optimum fungal pretreatment occurred in samples 3 and 4 for particle sizes of 200 μm in the wavelength region ranging between 2950 cm^{-1} to 2978 cm^{-1} since it reduces the size of individual fiber into the microfibrils size.

References

[1] Y. Habibi, L.A. Lucia and O.J. Rojas. (2010). Cellulose nanocrystals: Chemistry, self-assembly, and application. *Chemica Reviews*, 110 (6), pp. 3479-3500.

[2] D. N. Jimat, H. M. Salleh, and H. Anuar. (2016). Effect of Fungal Pretreatment by Aspergillus Niger Assisted by Ultrasonication on Microfibrillated Cellulose Obtained from Cocoa Pod Husk. *Akademia Baru*, 21(1), pp. 1–13.

[3] Tauler, R. *Chemometric and Intelligent Laboratory Systems*. https://www.journals.elsevier.com/chemometrics-and-intelligent-laboratory-systems

[4] R. M. Amir, F. M. Anjum, M. I. Khan, M. R. Khan, I. Pasha and M. Nadeem. (2013). Application of Fourier transform infrared (FTIR) spectroscopy for the identification of wheat varieties. *Journal Food Science Technology*, 50(5), pp. 1018-1023. [Online]. Available: http://doi: 10.1007/s13197-011-0424-y

[5] K. Banas, A. Banas, H. O. Moser, M. Bahou, M. Cholewa, P. Yang and S. K. Lim. (2010). Multivariate statistical methods in the forensic investigation of the post-blast residues measured by Fourier transform infrared spectroscopy. *Diamond Light Source Proceedings*, 1, pp. 7–10.

[Online]. Available: http://doi.org/10.1017/S2044820110000900.

[6]   J. S. Lupoi, S. Singh, B. A. Simmons and R. J. Henry. (2014). Assessment of Lignocellulosic
      Biomass Using Analytical Spectroscopy: an Evolution to High-Throughput Techniques.
      BioEnergy Resources, pp. 1–23. [Online]. Available: http://doi.org/10.1007/s12155-013-9352-.

[7]   A. K. Kumar and S. Sharma. (2017). Recent updates on different methods of pretreatment of
      lignocellulosic feedstocks: a review. *Bioresources and Bioprocessing*, 4. [Online]. Available:
      http://doi.org/10.1186/s40643-017-0137-9.