

Near infrared spectroscopy (NIRS) based high-throughput online assay for key cell wall features that determine sugarcane bagasse digestibility

Xinru Li

Guangxi University

Fumin Ma

Guangxi University

Chengping Liang

Guangxi University

Maoyao Wang

Guangxi University

Yan Zhang

Guangxi University

Yufei Shen

Guangxi University

Muhammad Adnan

Guangxi University

Pan Lu

Guangxi University

Muhammad Tahir Khan

Nuclear Institute of Agriculture

Jiangfeng Huang

Guangxi University <https://orcid.org/0000-0001-7784-4509>

Muqing Zhang (✉ zmuqing@163.com)

Nuclear Institute of Agriculture (NIA) <https://orcid.org/0000-0003-3138-3422>

Research

Keywords: Sugarcane, Cell wall, NIRS, Cellulose crystallinity, Lignin

Posted Date: October 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-90063/v1>

Abstract

Background: The improvement of sugarcane cell wall structure is a promising strategy to enhance the bagasse digestibility to improve its prospects as a bioenergy crop. In this context, cellulose crystallinity (CrI) and lignin are the key parameters that influence the saccharification efficiency. Therefore, this study was conducted to develop a high-throughput assay for online characterization of these cell wall features in sugarcane.

Results: A total of 838 different sugarcane genotypes were collected at different growth stages during 2018 and 2019. A continuous variation distribution of near-infrared spectroscopy (NIRS) was observed among the sugarcane samples. Due to significant diversity of the cell wall features in the sampled population of the crop, seven high quality calibration models were developed through online NIRS calibration. All of the generated equations displayed coefficient of determination (R^2) values higher than 0.8 and high ratio performance deviation (RPD) values over 2.0 in calibration, internal cross validation, and external validation. Particularly, the equations for CrI and the total lignin content exhibited the RPD values as high as 2.56 and 2.55, respectively, indicating their excellent prediction capacity. Furthermore, the offline NIRS assay was also performed. A comparable calibration was observed between the offline and online NIRS analyses, suggesting that both of the two strategies would be applicable for estimating cell wall characteristics. Nevertheless, as online NIRS assay offers greater advantages for large-scale screening jobs, it could be implied as a better option for high-throughput cell wall features prediction.

Conclusions: This study, as a foremost attempt, explored an online NIRS assay for high-throughput assessment of key sugarcane cell wall attributes in terms of CrI, lignin content, and its proportion in sugarcane. Consistent and precise calibration results were obtained in NIRS modeling; insinuating this strategy as a reliable approach for large-scale screening of promising sugarcane germplasm for cell wall structure improvement and beyond.

Background

Bioethanol has been recognized as a significant clean fuel for reducing carbon debt. In particular, cellulosic ethanol that is derived from lignocellulosic feedstock has received more attention because it does not compete with the food production or occupy the land otherwise used for this purpose [1]. In this regard, bagasse, a major byproduct of sugarcane crushing during juice extraction, shows great advantages for second-generation biofuel production [2]. However, due to cell wall recalcitrance to hydrolysis, the cost-effectiveness of cellulosic ethanol production from sugarcane remains a question [3]. Therefore, screening of germplasm for optimal cell wall features is important for using sugarcane as biofuel crop.

In general, plant cell wall is majorly composed of three different types of polymers i.e. cellulose, hemicellulose, and lignin. These polymers form a complex network structure that impedes cell wall digestibility [4]. Particularly, the properties of cellulose and lignin are closely related to the cell wall

recalcitrance [5–8]. or instance, cellulose is a polymer composed of glucose units linked *via* β -1,4-glycosidic bonds. Cellulose crystallinity index (CrI), that is characterized by X-rays scattering from crystalline and amorphous regions [9, 10], is a key parameter that defines the hindrance to the cell wall saccharification [5, 8, 11, 12]. While lignin is a hydrophobic polymer composed of phenylpropane compounds that often tightly associate with hemicellulose to form “LCC complex”. This “LCC complex” blocks the cellulose surface and hinders the cellulose accessibility [13]. Therefore, lignin is also a major factor affecting cell wall saccharification [6, 14, 15]. In this backdrop, the screening of germplasm resources for lower cellulose CrI and lignin content can play significant role for tweaking cell wall recalcitrance attributes. To carry-out large-scale screening studies for this purpose, a high-throughput assay for determining the cell wall characteristics is needed.

Near-infrared spectroscopy (NIRS) is a rapid and non-destructive analytical tool that is widely employed for high-throughput biomass quantity or quality analysis for biofuel production [16]. It has been used for characterization of cell wall polymer features [17–20], analysis of biomass saccharification efficiency [17, 18, 21], and prediction of ethanol production *via* yeast fermentation [22–25]. Notably, in sugarcane, some studies also have applied NIRS for determining cell wall components or prediction of digestibility [26–29]. In one of such efforts, Caliri et al. [30] explored the NIRS assay for estimating cellulose crystallinity index. However, most of such studies used offline calibration strategy that necessitates certain time-consuming steps for NIRS scanning. Hence, its application is limited in analyzing a large number of samples, generally required in crop improvement programs.

This study was initiated to develop a high-throughput online NIRS assay for characterizing key cell wall features in sugarcane bagasse. Hundreds of samples were collected from the germplasm population. Based on the standard laboratory analytical methods for cell wall features and the online system for near-infrared spectroscopy, a reliable online NIRS assay was developed for analyzing lignin content and CrI. Thus, this study provides a precise and high-throughput approach for large-scale screening and selection of optimal germplasm for reducing cell wall recalcitrance, to target genetic improvement of sugarcane for low-cost bioethanol production.

Results

Near-infrared spectroscopy-based characterization of sugarcane population

A total of 838 germplasm samples, brought in six different batches, were used for NIRS modeling. While analyzing each lot of samples, the NIRS data were immediately collected on an explicitly designed online system. The continuously collected spectrum reflectance values were averaged for NIRS calibration. As shown in Fig. 1A, the near infrared spectral reflectance values of all samples displayed a fluctuation within the normal range, indicating the diverse nature of these samples. To characterize the structure of the sampled population, principal component analysis (PCA) was carried out for the recorded near-infrared spectral values. In PCA analysis, new orthogonal variables were generated from the original

spectral values. The first thirteen principal components (PCs), which could explain 99.81% of the variation (Fig. 1B), were selected to characterize the population structure. A large variation of the sample population was observed within the selected PCs especially for the first five PCs (Fig. 1C). Finally, to take a view of the population structure, the first three PCs were used for a 3D observation of the sample distribution. As shown in Fig. 1D, although sugarcane samples of different genotypes were collected from different batches, no discriminable distribution was observed among them, suggesting that these samples could be exploited for a global NIRS modeling.

Diversity of cell wall features in collected sugarcane population

X-ray diffraction (XRD) method was applied for cellulose CrI determination. The maximum and the minimum diffraction were separately observed at the 2θ region ranging from 15° to 25° , allowing for a standard calculation of CrI. The varying diffraction values were observed in the sample population (Fig. 2A) depicting the diversity of the genotypes. The maximum and the minimum diffraction values were applied for cellulose CrI calculation. The CrI values of the sugarcane population were calculated to be ranging from 21–56% (Fig. 2B). These results were comparable with the previous reports for sugarcane and *Miscanthus* [8, 30]. The statistical distribution showed that cellulose CrI exhibited a normal distribution in the analyzed sugarcane population (Fig. 2C). The diversity of the CrI values indicated a large variation for cellulose-related features in the samples.

Lignin mass content (% dry mass) was analyzed through two-steps acid hydrolysis combined with ashing. Using this standard process, the acid soluble lignin (ASL), acid insoluble lignin (AIL), and the total of the two were determined. The ASL content (% dry mass) varied from 1.2% to 2.6%, while the AIL ranged from 9.2–25.3%. Moreover, a large variation of the total lignin content (% dry mass), which ranged from 10.9–27.0%, was observed (Fig. 2D & Table S1). The lignin content is closely related to the cell wall network structure that significantly impacts lignocellulose digestibility. Therefore, this study also estimated the lignin proportion in the sugarcane cell wall. The lignin proportion in the cell wall exhibited a greater variation than the one observed in case of clean dry mass, especially for the ASL values which illustrated the highest coefficient of variation (CV) of 0.19 (Fig. 2F & Table S1). The total lignin content (% cell wall) ranged from 24.3–56.2%, depicting a variation of the cell wall structure in collected sugarcane population. Furthermore, the frequency distribution analysis of lignin content in terms of clean dry mass and proportion was performed for determining the population structure. As a result, a normal distribution was observed for all these parameters under consideration (Fig. 2E & G), hinting a reliable NIRS calibration.

Characterization of calibration and validation sets

To take NIRS modeling and the accorded performance evaluation, the samples were divided into two sets, one was used for NIRS calibration and the other was applied for external validation. For cellulose CrI modeling, a total of 120 samples were randomly selected from sample population to build an external

validation set and the remaining 718 samples formed the calibration set (Fig. 3A). Similarly, a total of 679 samples were used for lignin content (% dry mass) modeling, 565 of which were used for calibration and 114 samples for external validation (Fig. 3B). For lignin proportion, 446 and 117 samples were analyzed for calibration and equation evaluation, respectively (Fig. 3C). Moreover, for comparison between calibration and validation sets, a frequency distribution was carried out for the cell wall features in both sets. Notably, all of them were comparable and showed a similar normal distribution (Fig. 3A, B & C). Hence, these comparable data sets allowed a reliable NIRS modeling and evaluation.

Online NIRS modeling

Partial least square (PLS) regression analysis method packed in OPUS software were used to perform NIRS modeling. Dozens of parameters were combined in terms of wavelength range selection and spectrum pretreatment to obtain calibration equations in PLS analysis. The internal cross validation or external validation were performed for evaluating the performance of the equations and then the best equations were attained (according to their high-performance in validation).

The calibration results showed that all of the equations produced for the two cell wall features exhibited high coefficient of determination (R^2) values *viz.* over 0.80, except for AIL proportion which showed an R^2 value of 0.78 (Fig. 4). The total lignin content equation in the dry biomass determination demonstrated the highest fit performance, recording the maximum R^2 value of 0.91 (Fig. 4B). Therefore, the identified excellent correlations between fit and the reference values during calibration indicated high prediction capacity of the obtained equations.

In addition, the samples from the external validation sets were applied for an independent validation assay to evaluate the prediction performance of the obtained models. A correlation analysis between the predicted values and the measured values was carried out, and root means square error of external validation (RMSECV) and ratio performance deviation (RPD) were additionally calculated. The results suggested that all of the equations exhibited high correlation between the predicted and the true values. The coefficient of determination of external validation (R^2_{ev}) ranged from 0.75 to 0.81 (Fig. 4). The AIL proportion showed R^2_{ev} value of 0.75 that was consistent with the calibration results. Notably, all of the equations obtained RPD values higher than 2.0 during external validation, suggesting their excellent prediction performance.

Finally, to achieve a better performance of the equations for cell wall features prediction, samples in the external validation set were combined into the calibration set for a global NIRS modeling. As more samples were added, wider variation of cell wall features was observed in the integrated new calibration sets (Table S2). As expected, most of the equations demonstrated substantial improvement in prediction capacity. In detail, the equation for cellulose CrI prediction showed the highest improvement as its R^2 value rose from 0.81 to 0.88 (Fig. 5A). For lignin content and proportion prediction, the AIL exhibited the maximum amelioration (Fig. 5B & 5C). All of the new equations obtained from this analysis exhibited high correlation between fit and the measured values proposing their excellent fitting during calibration. Furthermore, the internal cross validation assay was performed to evaluate the prediction capacity of the

equations. During cross validation, the calibration set was randomly partitioned into several groups and samples in each group were validated using a calibration equation developed on other samples. The results proposed that all of the generated equations exhibited high R^2_{cv} and RPD values; especially, the RPD values that ranged from 2.21 to 2.56 (Fig. 5D & 5C). The equations for lignin content (% dry mass) and cellulose Crl displayed the highest R^2_{cv} value of 0.85 (Fig. 5E), hinting a consistence with the calibration results. Notably, the AIL proportion illustrated a consistent and high R^2 and R^2_{cv} values of 0.80 in calibration and validation, illustrating their staunch prediction capacity (Fig. 5F). Taken together, all of the new equations demonstrated good R^2 , R^2_{cv} and RPD values of calibration and internal cross validation. Hence, the generated equations could be applied for determining the cell wall features.

Discussion

The plant cell wall structure governs the biomass digestibility. In particular, lignin is one of the key factors that hinder lignocellulose saccharification [4–7]. It is a hydrophobic polymer, which can adsorb cellulase and hinder the accessibility to cellulose [31, 32]. Thus, high lignin content indicates low cell wall digestibility. Moreover, lignin is composed of different monomers; the variations in these monomers dictate the lignin structure. Different monomers often show dissimilar cross-link patterns in lignin and hemicellulose interaction; and therefore, the lignin structure plays crucial role in defining the cell wall polymers network. For instance, syringyl/guaiacyl (S/G) ratio has been identified as one of the most important factors that negatively affects cell wall saccharification in *Switchgrass* [33]. Thus, the lignin structure seems to be a key role player that influences the lignocellulose digestibility. Keeping this background in view, in this study, six equations were generated for lignin content determination. Among them, three of the equations were developed for lignin mass content (% dry biomass) prediction; whereas, the other three were designed for characterizing the percentage of the lignin content (% cell wall). These specialized NIRS models could be applied for multi-purpose lignin content determination.

Cellulose is a polymer that can be saccharified for bioethanol production through fermentation. Cellulose Crl and the degree of polymerization (DP) are the main features that define the cellulose structure. Both of these traits are negatively correlated to saccharification efficiency and hinder the utilization of cellulose in second-generation ethanol production [8]. Interestingly, the samples with high DP generally exhibit high Crl as well and show low biomass digestibility [8, 12]. Therefore, any of these two features can be employed for cellulose structure characterization. In this study, although NIRS calibration was only carried out for cellulose Crl and lignin content, it could be applied for precise cellulose features prediction.

To reduce cell wall recalcitrance, attempts have been made to modify cell wall structure by reducing cellulose crystallinity and lignin content in sugarcane [34–37] and other energy plants [12, 38–41]. The transgenic plants engineered to aim desired variations in these characteristics have shown significant improvement of cell wall saccharification. Therefore, these cell wall features should be the traits of interest for energy cane breeding. The association studies through large-scale phenotypic and genotypic analyses have emerged as a promising strategy for crop improvement. However, due to the lack of the effective high-throughput phenotyping methods, it is difficult to obtain accurate phenotypic data.

Therefore, here, we report an online NIRS assay for high-throughput screening on the basis of cell wall features. The produced equations showed high $R^2/R^2_{cv}/R^2_{ev}$ values of calibration, internal cross validation, and external validation, suggesting their high-quality performance. Hence, they could be used for high-throughput phenotyping or screening of optimal germplasm for energy cane breeding.

The offline NIRS calibration was also conducted for cell wall features prediction. For this purpose, after online NIRS scanning, the shredded samples were collected, dried, and subjected to offline NIRS calibration. The results depicted some obvious differences of offline near infrared spectra vs. the online ones (Fig. 5A); yet, a continuous variation was evident (Fig. 5B). Similarly, partial least square (PLS) regression analysis was applied for offline NIRS modeling. In PLS analysis, high R^2 and RPD values of calibration were obtained for both cellulose CrI and lignin content. All of the equations exhibited a high linear correlation between the predicted and the reference values in internal cross validation (Fig. 5C-E). Notably, the offline NIRS models achieved better prediction performance than those reported previously [26, 30], which could be attributed to the large population of diverse samples employed for NIRS modeling in this study.

For precise evaluation of crop genotypes and reliable germplasm selection, samples should be analyzed as soon as possible after collection. The huge number of samples in such screening jobs necessitates the use of appropriate high-throughput techniques. The offline NIRS calibration uses ground dry samples. The additional steps of grinding and drying are time consuming and labor intensive and thus largely limit the use of offline NIRS strategies. Therefore, we carried out a comparison between offline and online analyses. Most of the online equations showed a comparable performance with the offline ones, while some of them even illustrated higher R^2 and RPD values of calibration and validation (Fig. 5 & Fig. 6). Thus, both of the two NIRS strategies could be used for determining the cell wall features. For large-scale screening jobs, the online NIRS assay is more advantageous and opportune. Hence, this strategy can be used as a convenient tool for screening of sugarcane germplasm.

Since lignin and cellulose features play critical role in cell wall recalcitrance, this study explored both the offline and online NIRS modeling for predicting these cell wall features. The future research may investigate the NIRS models for the lignin monomers contents and their ratios. All of equations produced in this research exhibited high prediction performance, suggesting their excellent potential for use in germplasm screening. Particularly, the online calibration models developed in this study, due to significant advantages in its protocols, exhibit excellent prospects for high-throughput screening of large-scale samples for energy cane breeding and germplasm selection.

Conclusions

This study developed an online NIRS assay for high-throughput analysis of key cell wall features in terms of cellulose crystallinity, lignin content, and its proportion in sugarcane. Consistent and precise calibration was obtained in NIRS modeling; suggesting this strategy as a reliable approach for large-scale screening

of optimal sugarcane germplasm. The results of the study can be employed for high-throughput screening jobs for bioenergy production from sugarcane.

Methods

Sample collection

A total of 838 sugarcane germplasm samples were planted at Fusui experimental field of Guangxi University. Hundreds of them were harvested at different growth stages in the year 2018 and 2019. In detail, 164, 162, 184, 70, and 48 samples were collected once a month from November 2018 to March 2019; moreover, 210 samples were collected in December 2019. Six stalks were randomly selected for each genotype and further analysis was done after removing leaves and young tips.

Near-infrared spectral data collection

The selected six stalks of each genotype were shredded using DM540 (IRBI Machines & Equipment Ltd, Brazil). The shredded fresh samples were immediately blended and transferred for NIRS scanning by CPS (Cane presentation system, Bruker Optik GmbH, Germany). Near infrared spectral data of fresh samples were simultaneously collected through MATRIX-F (Bruker Optik GmbH, Germany) online system. Following the online NIR spectral data collection, the shredded samples were dried under 60°C and then at 100 °C for 1 h after inactivation. The sample was ground through a 40-mesh screen and stored in a dry container until use. MATRIX-F equipped with Q413 sensor head was used for offline NIR spectral data collection.

Full band scanning mode with the wavelengths ranging from 4000 to 10000 cm^{-1} with 4 cm^{-1} steps was employed for collecting on-line and offline spectral data. The spectral absorbance values were recorded as $\log 1/R$, where R is the sample reflectance. The obtained reflectance values were then averaged for further analysis.

Lignin content determination

Two-step acid hydrolysis was used for determining the lignin content according to the analytical procedure of the National Renewable Energy Laboratory [42]. Briefly, 0.50 g of ground dry samples were extracted using benzene-ethanol (2:1, v/v) in a Soxhlet for 4 h and then hydrolyzed using 10.0 mL 67% (v/v) H_2SO_4 (at 25 °C for 90 min with a gentle shaking at 115 RPM). After hydrolysis, the acid solution was subsequently diluted to 3.97% (w/w) with distilled water and heated at 115 °C for 60 min. The autoclaved hydrolysis solution was filtered through a filtering crucible. The supernatant liquids were fixed to 250 mL and read at 205 nm under UV spectroscopy to estimate acid soluble lignin. The remaining residues were ashed in a muffle furnace at $575 \text{ °C} \pm 25 \text{ °C}$ for 4 h to ascertain the acid insoluble lignin. All experiments were conducted in triplicate.

Lignocellulose crystallinity index determination

X-ray diffraction (XRD) method was used to determine lignocellulose crystallinity index (CrI) as described by Zhang et al. 2013 [8]. In detail, approximately 0.3 g of the ground dry samples were extracted using 10 mL of distilled water to remove the soluble sugar. The residues were subsequently extracted using chloroform-methanol (1:1, v/v), methanol, and acetone, and then dried under vacuum conditions. The remaining residues were classified as crude cell walls and were used for examining through XRD.

Rigaku-D/MAX 2500V instrument (Ultima III, Japan) was employed for XRD analysis. The crude cell wall powder was laid on the glass holder and investigated under plateau conditions. For this analysis Ni-filtered Cu-K α radiation ($\lambda = 0.154056$ nm) generated at 40 KV voltage and 18 mA was used. A continuous scanning from 10 to 45° was performed at the speed of 0.0197°/s. The CrI was estimated in terms of percentage by calculating the intensity of the 200 peak (I_{200} , $h = 22.5^\circ$) and the intensity at the minimum between the 200 and 110 peaks (I_{am} , $h = 18.5^\circ$) as follows: $CrI = 100 \times (I_{200} - I_{am}) / I_{200}$, where I_{200} represents both crystalline and amorphous materials while I_{am} denotes amorphous material [43].

NIRS data processing and calibration

The OPUS spectroscopy software (version 7.8, Bruker Optik GmbH, Germany) was used for data processing and NIRS calibration. To solve the problems associated with the overlapping peaks and baseline correction, pretreatment and the wavelength range selection of the raw spectral data was performed before calibration. Several spectral pretreatment methods were used in OPUS software, namely constant offset elimination (COE), straight line subtraction (SSL), standard normal variate (SNV), Min-Max normalization (MMN), multiplicative scattering correction (MSC), first derivative (FD), second derivative (SED), a combinations of first derivative and straight line subtraction (FD + SSL), a combinations of first derivative and standard normal variate (FD + SNV), and a combination of first derivative and multiplicative scattering correction (FD + MSC). The NIRS spectra was divided into multiple intervals and then reassembled to obtain the optimal spectral region for calibration. A principal component analysis (PCA) was carried out to characterize the structure of spectral population and the GH outlier (GH > 3.0) samples were eliminated. Moreover, partial least square (PLS) regression was performed to produce calibration equations. The internal cross validation and external validation were carried out to test the performance of the generated equations. The best equations were selected according to the high coefficient of determination of the calibration/internal cross validation/external validation ($R^2/R^2_{cv}/R^2_{ev}$), low root mean square error of calibration/internal cross validation/external validation (RMSEC/RMSECV/RMSEP), and high ratio performance deviation (RPD) values [17, 21].

Declarations

Authors' contributions

XL and FM completed major experiment, analyzed the data. CL, MW, YZ, YS, MA and PL participated in the determination of lignin content and lignocellulose crystallinity index analysis. JH and MZ designed

the project, supervised the experiments, interpreted the data, and finalized the manuscript. MTK revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Consent for publication

Not applicable.

Funding

This work was funded by the National Natural Science Foundation on China (31660420); the Earmarked Fund for Modern Agro-industry Technology Research System“CARS170-190”; Science and Technology Major Project of Guangxi (AA17202042-7 and Gui Ke 2018-266-Z01); the State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources (SKLCUSA-a202002).

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its Additional files.

Ethics approval and consent to participate

Not applicable.

References

1. Somerville C, Youngs H, Taylor C, Davis SC, Long SP. Feedstocks for Lignocellulosic Biofuels. *Science*. 2010;329:790–2.
2. Huang J, Khan MT, Perecin D, Coelho ST, Zhang M. Sugarcane for bioethanol production: Potential of bagasse in Chinese perspective. *Renewable and Sustainable Energy Reviews*. 2020; 133.

3. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*. 2007;315:804–7.
4. Zhao X, Zhang L, Liu D. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. *Biofuels Bioproducts Biorefining-Biofpr*. 2012;6:465–82.
5. Chang VS, Holtzapple MT. Fundamental factors affecting biomass enzymatic reactivity. *Applied Biochemistry and Biotechnology*. 2000; 84 – 6:5–37.
6. Si SL, Chen Y, Fan CF, Hu HZ, Li Y, Huang JF, Liao HF, Hao B, Li Q, Peng LC, Tu YY. Lignin extraction distinctively enhances biomass enzymatic saccharification in hemicelluloses-rich *Miscanthus* species under various alkali and acid pretreatments. *Biores Technol*. 2015;183:248–54.
7. dos Santos AC, Ximenes E, Kim Y, Ladisch MR. Lignin-Enzyme Interactions in the Hydrolysis of Lignocellulosic Biomass. *Trends Biotechnol*. 2019;37:518–31.
8. Zhang W, Yi Z, Huang J, Li F, Hao B, Li M, Hong S, Lv Y, Sun W, Ragauskas A, et al. Three lignocellulose features that distinctively affect biomass enzymatic digestibility under NaOH and H₂SO₄ pretreatments in *Miscanthus*. *Bioresour Technol*. 2013;130:30–7.
9. Thygesen A, Oddershede J, Lilholt H, Thomsen AB, Stahl K. On the determination of crystallinity and cellulose content in plant fibres. *Cellulose*. 2005;12:563–76.
10. Bansal P, Hall M, Realff MJ, Lee JH, Bommarius AS. Multivariate statistical analysis of X-ray data from cellulose: A new method to determine degree of crystallinity and predict hydrolysis rates. *Biores Technol*. 2010;101:4461–71.
11. Laureano-Perez L, Teymouri F, Alizadeh H, Dale B. Understanding factors that limit enzymatic hydrolysis of biomass: characterization of pretreated corn stover. *Appl Biochem Biotechnol*. 2005;121–124:1081–99.
12. Huang JF, Xia T, Li GH, Li XL, Li Y, Wang YT, Wang YM, Chen YY, Xie GS, Bai FW, et al. Overproduction of native endo- β -1,4-glucanases leads to largely enhanced biomass saccharification and bioethanol production by specific modification of cellulose features in transgenic rice. *Biotechnol Biofuels*. 2019; 12.
13. Zhao Y, Shakeel U, Rehman MSU, Li HQ, Xu X, Xu J. Lignin-carbohydrate complexes (LCCs) and its role in biorefinery. *Journal of Cleaner Production*. 2020; 253.
14. Johnston PA, Zhou H, Aui A, Wright MM, Wen Z, Brown RC. A lignin -first strategy to recover hydroxycinnamic acids and improve cellulosic ethanol production from corn stover. *Biomass & Bioenergy*. 2020; 138.
15. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M, et al. Lignin Valorization: Improving Lignin Processing in the Biorefinery. *Science*. 2014;344:1246843.
16. Xu F, Yu J, Tesso T, Dowell F, Wang D. Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. *Appl Energy*. 2013;104:801–9.

17. Huang J, Li Y, Wang Y, Chen Y, Liu M, Wang Y, Zhang R, Zhou S, Li J, Tu Y, et al. A precise and consistent assay for major wall polymer features that distinctively determine biomass saccharification in transgenic rice by near-infrared spectroscopy. *Biotechnol Biofuels*. 2017; 10.
18. Wu L, Li M, Huang J, Zhang H, Zou W, Hu S, Li Y, Fan C, Zhang R, Jing H, et al. A near infrared spectroscopic assay for stalk soluble sugars, bagasse enzymatic saccharification and wall polymers in sweet sorghum. *Biores Technol*. 2015;177:118–24.
19. Milano ER, Payne CE, Wolfrum E, Lovell J, Jenkins J, Schmutz J, Juenger TE. Quantitative trait loci for cell wall composition traits measured using near-infrared spectroscopy in the model C4 perennial grass *Panicum hallii*. *Biotechnology for Biofuels*. 2018; 11.
20. Payne CE, Wolfrum EJ. Rapid analysis of composition and reactivity in cellulosic biomass feedstocks with near-infrared spectroscopy. *Biotechnology for Biofuels*. 2015; 8.
21. Huang J, Xia T, Li A, Yu B, Li Q, Tu Y, Zhang W, Yi Z, Peng L. A rapid and consistent near infrared spectroscopic assay for biomass enzymatic digestibility upon various physical and chemical pretreatments in *Miscanthus*. *Bioresour Technol*. 2012;121:274–81.
22. Vogel KP, Dien BS, Jung HG, Casler MD, Masterson SD, Mitchell RB. Quantifying Actual and Theoretical Ethanol Yields for Switchgrass Strains Using NIRS Analyses. *Bioenergy Res*. 2011;4:96–110.
23. Schmer MR, Vogel KP, Mitchell RB, Dien BS, Jung HG, Casler MD. Temporal and Spatial Variation in Switchgrass Biomass Composition and Theoretical Ethanol Yield. *Agronomy Journal*. 2012;104:54–64.
24. Guimaraes CC, Simeone MLF, Parrella RAC, Sena MM. Use of NIRS to predict composition and bioethanol yield from cell wall structural components of sweet sorghum biomass. *Microchemical Journal*. 2014;117:194–201.
25. Li M, Wang J, Du F, Diallo B, Xie GH. High-throughput analysis of chemical components and theoretical ethanol yield of dedicated bioenergy sorghum using dual-optimized partial least squares calibration models. *Biotechnology for Biofuels*. 2017; 10.
26. Assis C, Ramos RS, Silva LA, Kist V, Barbosa MHP, Teofilo RF. Prediction of Lignin Content in Different Parts of Sugarcane Using Near-Infrared Spectroscopy (NIR), Ordered Predictors Selection (OPS), and Partial Least Squares (PLS). *Appl Spectrosc*. 2017;71:2001–12.
27. Sabatier D, Thuries L, Bastianelli D, Dardenne P. Rapid prediction of the lignocellulosic compounds of sugarcane biomass by near infrared reflectance spectroscopy: comparing classical and independent cross-validation. *Journal of near Infrared Spectroscopy*. 2012;20:371–85.
28. Chong BF, O'Shea MG. Advancing Energy Cane Cell Wall Digestibility Screening by Near-Infrared Spectroscopy. *Appl Spectrosc*. 2013;67:1160–4.
29. Hoang NV, Furtado A, Donnan L, Keefe EC, Botha FC, Henry RJ. High-Throughput Profiling of the Fiber and Sugar Composition of Sugarcane Biomass. *Bioenergy Res*. 2017;10:400–16.
30. Caliarì IP, Barbosa MHP, Ferreira SO, Teofilo RF. Estimation of cellulose crystallinity of sugarcane biomass using near infrared spectroscopy and multivariate analysis methods. *Carbohydr Polym*.

2017;158:20–8.

31. Li M, Yi L, Bin L, Zhang Q, Song J, Jiang H, Chen C, Wang S, Min D. Comparison of nonproductive adsorption of cellulase onto lignin isolated from pretreated lignocellulose. *Cellulose*. 2020;27:7911–27.
32. Jin W, Chen L, Hu M, Sun D, Li A, Li Y, Hu Z, Zhou S, Tu Y, Xia T, et al. Tween-80 is effective for enhancing steam-exploded biomass enzymatic saccharification and ethanol production by specifically lessening cellulase absorption with lignin in common reed. *Appl Energy*. 2016;175:82–90.
33. Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M Jr, Chen F, Foston M, Ragauskas A, Bouton J, et al. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc Natl Acad Sci USA*. 2011;108:3803–8.
34. Jung JH, Fouad WM, Vermerris W, Gallo M, Altpeter F. RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass. *Plant Biotechnol J*. 2012;10:1067–76.
35. Jung JH, Vermerris W, Gallo M, Fedenko JR, Erickson JE, Altpeter F. RNA interference suppression of lignin biosynthesis increases fermentable sugar yields for biofuel production from field-grown sugarcane. *Plant Biotechnol J*. 2013;11:709–16.
36. Jung JH, Kannan B, Dermawan H, Moxley GW, Altpeter F. Precision breeding for RNAi suppression of a major 4-coumarate:coenzyme A ligase gene improves cell wall saccharification from field grown sugarcane. *Plant Mol Biol*. 2016;92:505–17.
37. Kannan B, Jung JH, Moxley GW, Lee SM, Altpeter F. TALEN-mediated targeted mutagenesis of more than 100 COMT copies/alleles in highly polyploid sugarcane improves saccharification efficiency without compromising biomass yield. *Plant Biotechnol J*. 2018;16:856–66.
38. Li Y, Liu P, Huang JF, Zhang R, Hu Z, Feng SQ, Wang YT, Wang LQ, Xia T, Peng LC. Mild chemical pretreatments are sufficient for bioethanol production in transgenic rice straws overproducing glucosidase. *Green Chem*. 2018;20:2047–56.
39. Chen F, Dixon RA. Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol*. 2007;25:759–61.
40. Eudes A, Sathitsuksanoh N, Baidoo EEK, George A, Liang Y, Yang F, Singh S, Keasling JD, Simmons BA, Loque D. Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. *Plant Biotechnology Journal*. 2015;13:1241–50.
41. Zhang J, Li M, Bryan AC, Yoo CG, Rottmann W, Winkeler KA, Collins CM, Singan V, Lindquist EA, Jawdy SS, et al. Overexpression of a serine hydroxymethyltransferase increases biomass production and reduces recalcitrance in the bioenergy crop *Populus*. *Sustainable Energy Fuels*. 2019;3:195–207.
42. Sluiter A, Hames B, Scarlata C, Sluiter J, Templeton D, Crocker D. Determination of structural carbohydrates and lignin in biomass national renewable. *Energy Lab*. 2011;10:1–15.
43. Segal L, Creely J, Martin A, Conrad C. An empirical method for estimating the degree of crystallinity of native cellulose using the X-Ray diffractometer. *Textile Research Journal*. 1959;29:786–94.

Figures

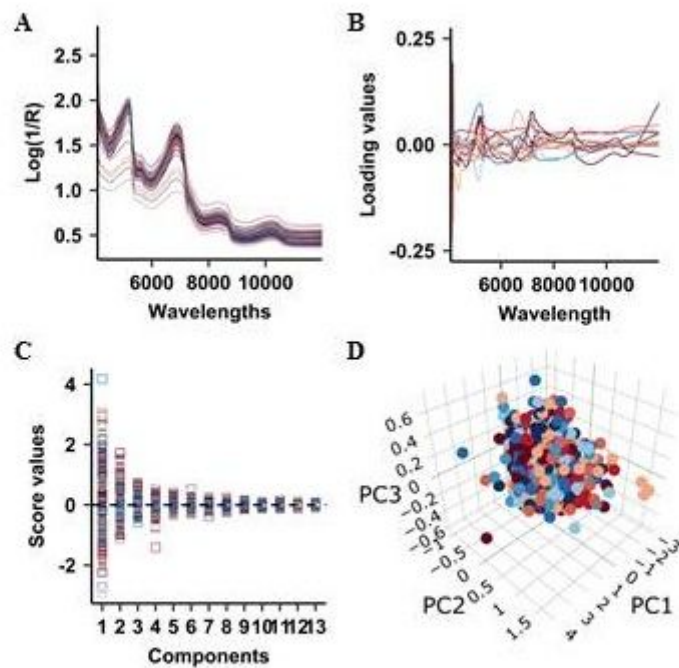


Fig. 1

Figure 1

Near-infrared spectra characterization in sugarcane population. (A) Original spectroscopy. (B) The first 13 principal components for population characterization. (C) Sample variations on each principal component in sugarcane population. (D) 3D view of sugarcane population structure.

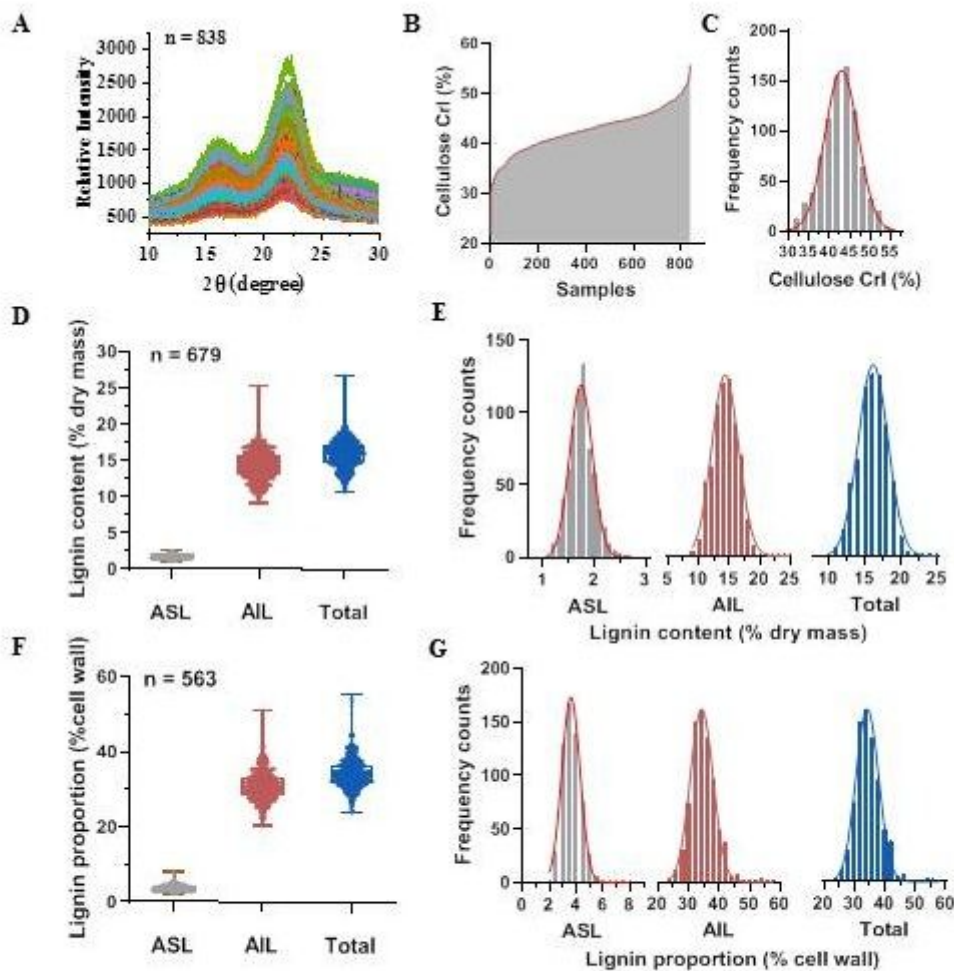


Fig. 2

Figure 2

Diversity of cell wall features in sugarcane population. (A) X-ray diffractograms. (B) Variation of cellulose crystallinity. (C) Cellulose crystallinity distribution in sugarcane population. (D-E) Variation of lignin clean mass content (D) and distribution (E) in sugarcane population. (F-G) Variation of lignin proportion in cell wall (F) and distribution (G) in sugarcane population.

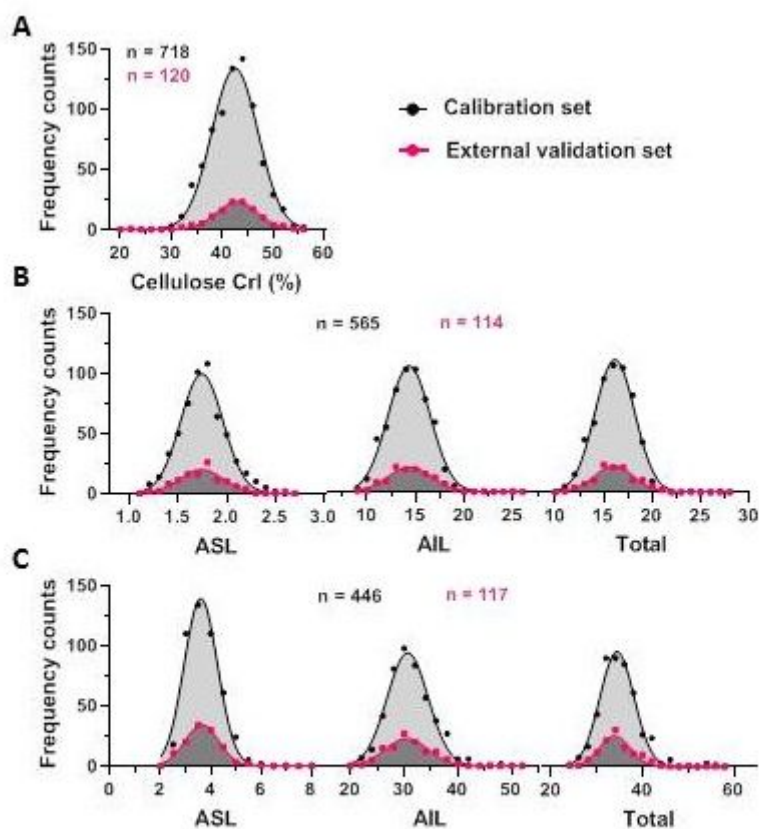


Fig. 3

Figure 3

Samples distribution in calibration and validation set for online NIRS modeling. (A) Cellulose crystallinity. (B) Lignin clean mass content in dry biomass. (C) Lignin proportion in cell wall.

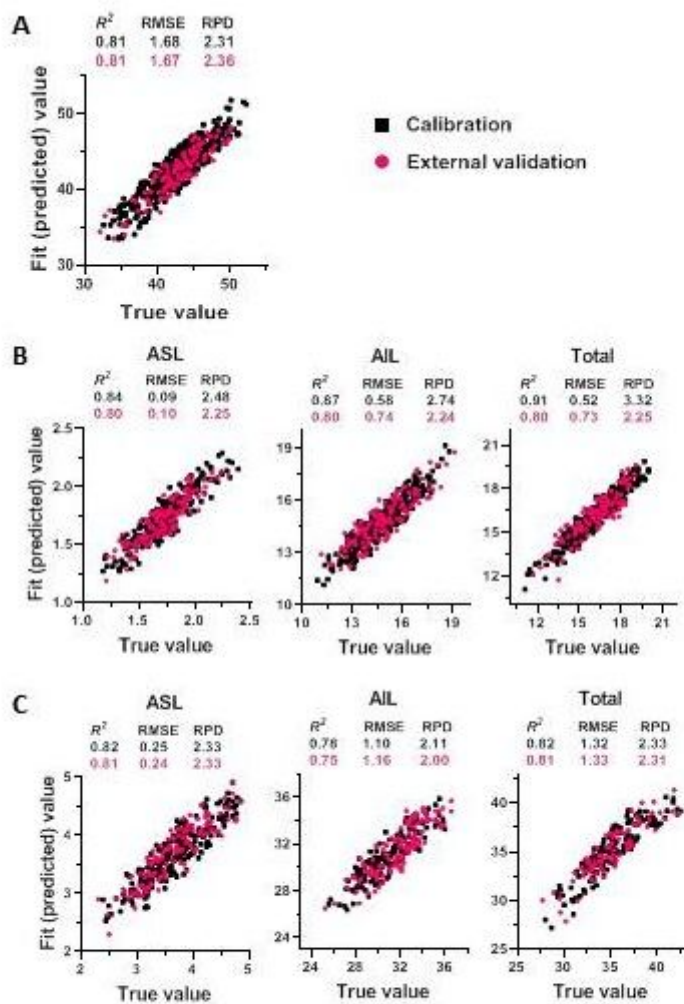


Fig. 4

Figure 4

Equations' performance during online NIRS calibration and external validation. (A) Cellulose crystallinity. (B) Lignin clean mass content in dry biomass. (C) Lignin proportion in cell wall.

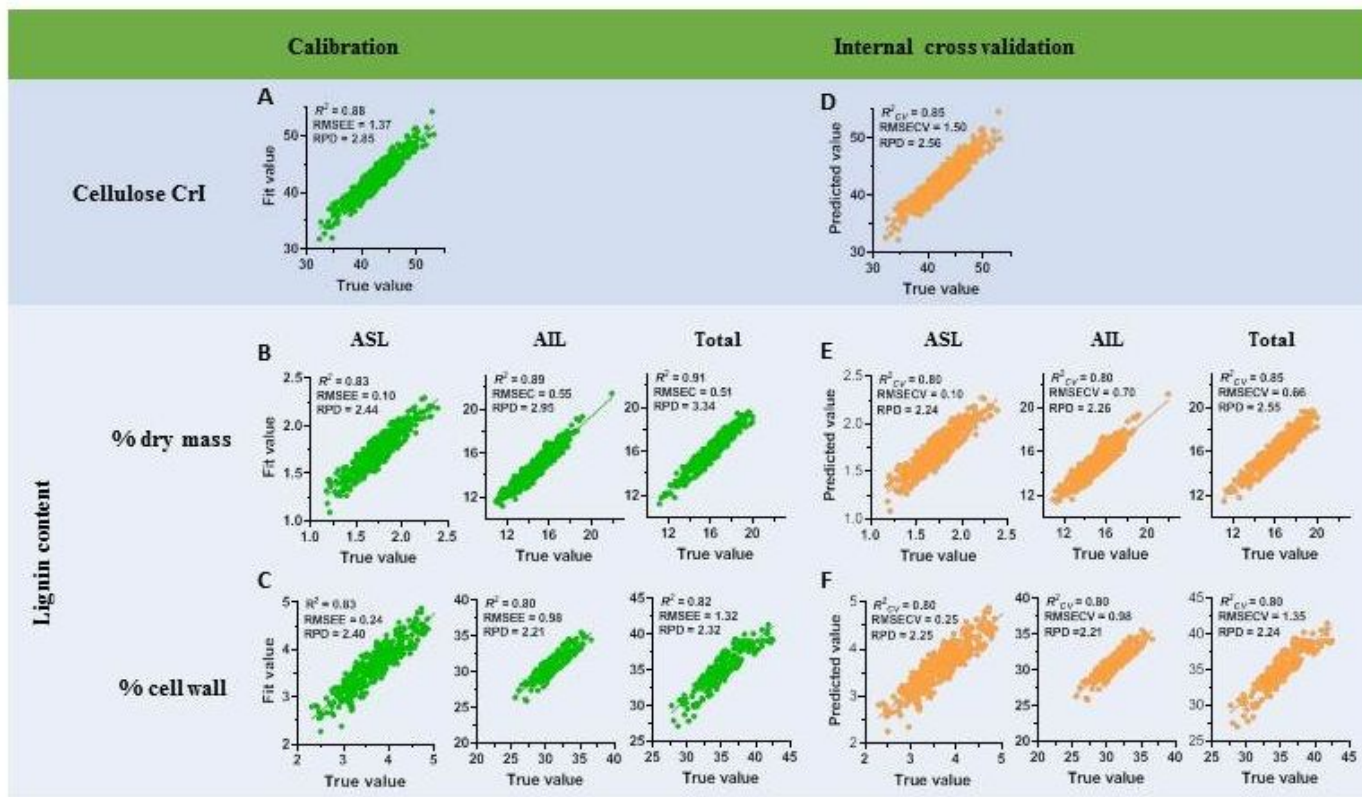


Fig. 5

Figure 5

Equations' prediction performance in integrative online modeling. A-C: Calibration for (A) Cellulose crystallinity; (B) Lignin clean mass content in dry biomass; (C) Lignin proportion in cell wall. D-F: Internal cross validation for (D) Cellulose crystallinity; (E) Lignin clean mass content in dry biomass; (F) Lignin proportion in cell wall.

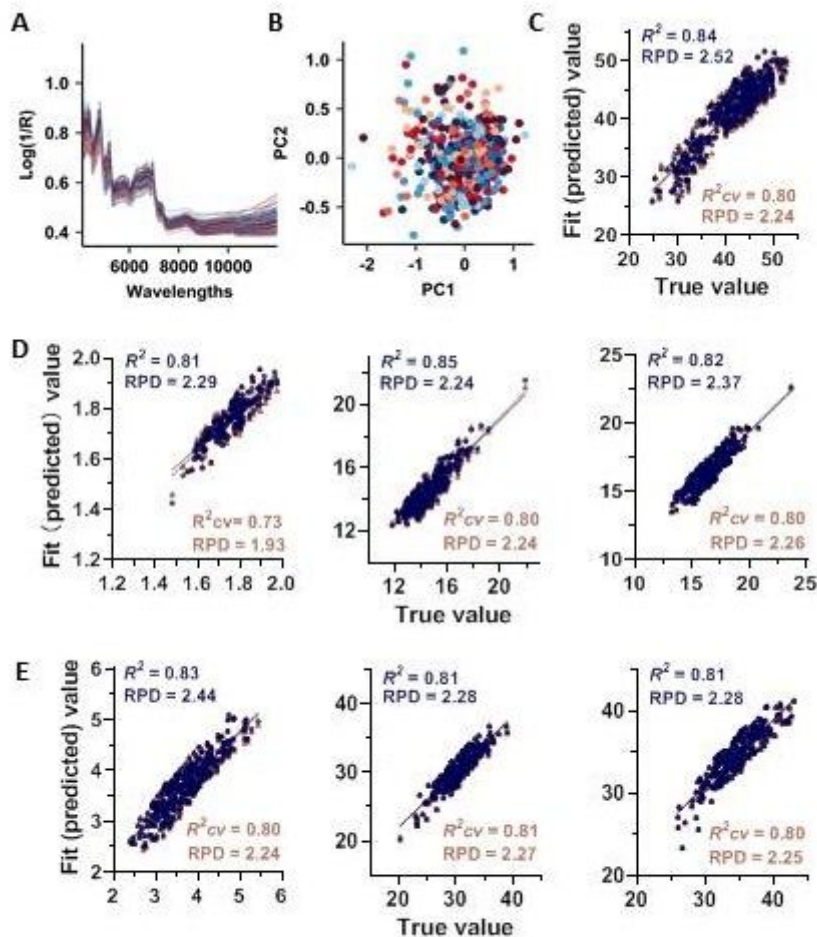


Fig. 6

Figure 6

Equations' prediction performance of offline NIRS models. (A) Original spectroscopy of sugarcane dry samples. (B) Principal components analysis of samples in calibration sets. C-E: Calibration model of (C) Cellulose crystallinity; (D) Lignin clean mass content in dry biomass; (E) Lignin proportion in cell wall.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.pptx](#)