Analysis of miscanthus biomass for biofuel production

Miguel Isauro Santillán Martínez

MSc. Minor Thesis Report
Analysis of miscanthus biomass for biofuel production

MSc. Minor Thesis Report

Submitted to the Department of Plant Breeding of Wageningen University in partial fulfilment for the requirements for the degree of

Master in Plant Biotechnology

With specialisation

Molecular Plant Breeding and Pathology

by

Miguel Isauro Santillán Martínez
870724-727-080

PBR-80424

Supervisors:
Luisa Trindade
Tim van der Weijde

Wageningen University and Research Centre

January, 2015
Abstract

Second generation bioethanol is one the most promising options for the production of sustainable and renewable liquid fuels. It makes use of the largely unutilized saccharides present in the plant cell walls. Miscanthus is an ideal prospect for the production of feedstock for bioethanol due to its high-yielding and nutrient-efficient characteristics. The production of bioethanol from lignocellulosic material still faces challenges that need to be overcome in order to achieve commercial production. One of the most evident areas of improvement in this process is the pre-treatment step, where the feedstock is thermo-chemically treated in order to maximise the release of fermentable sugars from the cell walls. Optimizing the cell wall composition by means of plant breeding can help to reduce the chemical intensity of the pre-treatments, and ultimately, reduce the costs and increase the efficiency of the production of second-generation biofuels. In this study, additional to cell wall composition, plants from the second year of OPTIMISC multi-location trials were analysed for saccharification efficiency. Glucose conversion was predicted with the use of near-infrared spectroscopy (NIRS) based on a model developed with data generated in the previous year. Additionally, a new cross-validated model also yielded good predictions of this trait. The results of this thesis will help to improve the prediction models for cell wall composition and saccharification efficiency and provide interesting insights on the variation on these traits in miscanthus plants on the first years after establishment.
## Contents

1. Introduction ........................................................................................................................................... 1
   1.1 Biofuels ................................................................................................................................................. 1
   1.2 Second generation bioethanol ................................................................................................................... 1
   1.3 Miscanthus .............................................................................................................................................. 2
   1.4 Biomass quality in biofuel production: the importance of cell wall composition ................................. 3
   1.5 OPTIMISC ........................................................................................................................................... 4
   1.6 Near-infrared spectroscopy (NIRS) ......................................................................................................... 6
   1.7 Thesis objectives ..................................................................................................................................... 8

2. Materials and methods ............................................................................................................................... 9
   2.1 Plant material ......................................................................................................................................... 9
   2.2 Spectral data acquisition. ......................................................................................................................... 9
   2.3 Sample selection for calibration and validation. ....................................................................................... 9
   2.4 Dry matter content (DMC) and neutral detergent fibre (NDF) analysis .............................................. 9
   2.5 Saccharification analysis ......................................................................................................................... 10
   2.6 Data analysis ......................................................................................................................................... 11
   2.7 Selection of the prediction model ........................................................................................................... 12

3. Results and discussion ............................................................................................................................... 13
   3.1 Prediction of dry matter content and neutral detergent fibre ................................................................. 13
   3.2 Determination of glucose conversion for calibration and validation sets ............................................. 13
   3.3 NIRS calibration models for saccharification efficiency ......................................................................... 15
   3.4 Prediction of saccharification efficiency ................................................................................................. 16

4. Conclusions and future prospects ............................................................................................................. 22

5. References .................................................................................................................................................. 23

Acknowledgements ................................................................................................................................. 25
1. Introduction

1.1 Biofuels

In recent decades, the increasing demand of fossil fuels, together with the adverse environmental effects associated with their production and use have urged the development of new and more sustainable sources of energy. Biofuels have arisen as an attractive and promising alternative of renewable energy. The production processes to obtain these biofuels serves to classify them into: first, second and third –generation biofuels (Demirbas, 2011).

First-generation biofuels are derived from sugars and oils from grains and seeds (Nigam&Singh, 2011). Three main types of first-generation biofuels are produced commercially, namely bioethanol, biodiesel and biogas (Naik et al., 2010). Out of these, ethanol produced from maize starch and sugar cane sucrose represents the biggest portion of biofuel production worldwide, concentrated in the United States and Brazil (Havlik et al., 2011). However, the production of fuel from food crops comes with large societal, environmental and economical concerns (Nigam&Singh, 2011), largely because of their negative impact on food security, extensive use of productive land and generally intensive use of resources, such as fertilization and irrigation (Mohr&Raman, 2013).

Second-generation biofuels are produced from lignocellulosic material obtained from the cell wall of plants, which comprises the majority of all plant biomass (Nigam&Singh, 2011). This is their main advantage compared to first-generation biofuels, as the production of feedstock does not necessarily compete with that of food crops. Additionally, they represent a more sustainable alternative to first-generation biofuels, as the impact of their production, in terms of CO₂ concentrations, can achieve neutral or even negative levels (Naik et al., 2010). However, various processes for generating second-generation biofuels still face technical challenges that need to be overcome before commercial production is achieved.

Third generation biofuels are produced from algae and the technology and algal strains to produce them in a sustainable manner are still being under development and investigation (Li-Beisson&Peltier, 2013).

1.2 Second generation bioethanol

The production of bioethanol from lignocellulosic material is one of the most promising options in the field of second-generation biofuels. Second generation bioethanol is produced by the fermentation of sugars obtained from the cell wall of plants. The amount of sugars present in these cell walls is very high and it is a largely unutilized source of saccharides. Furthermore, the production
of cell wall in the plant is naturally energy-efficient (Pauly&Keegstra, 2008). Vegetal cell walls comprise most of the total plant biomass and around 75% of the cell wall portion is composed of polysaccharides (Naik et al., 2010). Yet, several steps are required to obtain and process these sugars into bioethanol. These steps include biomass harvest, physical and thermo-chemical pre-treatments, enzymatic saccharification, fermentation and separation (Douglas et al., 2012).

Pre-treatments include a variety of mechanical, thermal and/or chemical procedures, whose objectives are to maximize exposure of the chemical bonds of the biomass to enzymes that mediate the hydrolysis of sugars (Douglas et al., 2012). After harvested and collected, a physical pre-treatment is applied to the lignocellulosic feedstock in order to reduce its particle size and the cellulose crystallinity. The main goals of this process is to increase the specific surface and reduce the degree of polymerization of the material, which ultimately result in a better digestibility. This process can be done by chopping, grinding and milling the biomass (Sun&Cheng, 2002). Afterwards, thermo-chemical pre-treatments are needed to separate the polysaccharides present in the lignocellulosic feedstock (Pauly&Keegstra, 2010). The composition of the cell wall of the feedstock highly influence the energy cost of these pre-treatments (Section 1.4 of this thesis).

Enzymatic saccharification, also referred as enzymatic hydrolysis, is needed to deconstruct cell wall polysaccharides of the biomass into their fermentable monosaccharide building blocks (Naik et al., 2010). The enzymes used in this process need to display superior hydrolysis kinetics on natural substrates and good stability properties (Douglas et al., 2012). Once monosaccharides are released, yeast fermentation and further distillation are used to obtain bioethanol. Saccharification and fermentation can be made simultaneously or separately (Douglas et al., 2012).

1.3 Miscanthus

One of the most promising crops for second-generation biofuels and other bio-based products is the grass miscanthus. Important features of this crop make it an ideal prospect for production of lignocellulosic feedstock (Clifton-Brown et al., 2008). Miscanthus is a high-yielding crop: under optimal conditions, the total yield can reach up to 45 Mg/ha (Heaton et al., 2010). It is also a perennial C₄ plant, which makes it a high resource-use efficient plant. Furthermore, it is a rhizomatous plant, which means that at the end of each growing season, the nutrients are translocated from the above-ground biomass to the below-ground rhizomes, reducing the amount of fertilizer to be used for the next season. Miscanthus is also broadly adapted to a variety of soils and climates (Heaton et al., 2010).

The genus Miscanthus includes around 17 species of perennial, non-wood rhizomatous grasses originating from Asia (Brosse et al., 2012). Miscanthus x giganteus is one of the highest yielding crop
of the genus. This genotype is an allotriploid hybrid originating from a cross between *M. sinensis* and *M. sacchariflorus*. Unfortunately, it is an sterile crop that can only be clonally propagated (Heaton *et al.*, 2010). This constitutes a challenge in establishing plant breeding programmes for this crop. However, efforts towards genetic improvement of miscanthus are underway. In the last few years, varieties with high potential for genetic improvement have been identified. Particularly, *M. sinensis* has a rich genetic diversity and have evolved populations that are adapted to a broad range of climates (Brosse *et al.*, 2012). In order to improve the productivity and range of adaptation, the genetic diversity of the *Miscanthus* genus needs to be exploited. Most breeding efforts to improve miscanthus have targeted mainly yield related traits, among others, drought, frost and low temperature tolerance, flowering time and propagation (Clifton-Brown *et al.*, 2008).

### 1.4 Biomass quality in biofuel production: the importance of cell wall composition

Cell walls of plants are comprised by three layers, namely middle lamella, primary cell wall and secondary cell wall (Buchanan *et al.*, 2006). The middle lamella constitutes an outermost layer that maintains the cell walls adhered to each other. The primary cell wall is flexible and is formed as the cell is growing. The secondary cell wall is a thick layer formed in the inner side of the primary cell wall.

Recent studies have focused on the importance of the composition of the cell wall of miscanthus and their impact on the optimization on the biofuel production, especially on the thermo-chemical pre-treatment step (Arnoult&Brancourt-Hulmel, 2014). These studies have identified significant variation across the different genotypes. It has been shown that cell wall of miscanthus is composed mainly by cellulose (40 to 60%), hemicellulose (20 to 40%) and lignin (10 to 30%) (Brosse *et al.*, 2012). Although the specific requirements for optimizing the lignocellulosic biomass are currently still being studied and defined, the generally desired characteristics of feedstock for bioethanol production are a high content of cellulose and hemicellulose and low content of lignin (Arnoult&Brancourt-Hulmel, 2014).

According to their polysaccharide composition, plant cell walls can be classified as type I or type II (Pauly&Keegstra, 2010). Miscanthus cell walls belong to the type II category, containing arabinoxylan as the major hemicellulose (cell walls belonging to type I contain xyloglucan as main hemicellulose). A key characteristic of type II cell walls is that lignin is highly accumulated in the secondary wall. This accumulation of lignin has a particularly important impact on the efficiency of the pre-treatments to release the fermentable sugars from the feedstock, as is one of the main factors contributing to the recalcitrance to degradation.
All plant cell walls are naturally recalcitrant to biological degradation and thus, energy-intensive conditions have to be applied to release the fermentable sugars from lignocellulosic feedstock. The natural factors contributing to this recalcitrance include (Himmel et al., 2007):

1. The cuticle and epicuticular waxes that constitute the epidermal tissue of the plant.
2. The presence of vascular bundles, their arrangement and density.
3. The amount of thick wall cell tissue (sclerenchymatous tissue).
4. The presence of lignin.
5. The high complexity of the cell wall constituents.
6. The inaccessibility of the cell wall component to the acting enzymes.
7. The presence and generation of inhibitors of further fermentations in the cell wall structure.

The cell wall composition of the feedstock thus have an enormous impact on the overall efficiency of the bioethanol production. Recently, it has been shown that maize genotypes with low concentration of cell wall lignin and high contents of stem cellulose and highly substituted hemicellulloses have a more effective bioconversion and glucose yields in mild dilute-acid pre-treatments (Torres et al., 2013). This strongly suggests that optimizing the cell wall composition by means of plant breeding can help reduce the chemical intensity of the pre-treatments, and ultimately, reduce the costs and increase the efficiency of the production of second-generation biofuels.

1.5 OPTIMISC

OPTIMISC is an international collaborative effort that aims for the optimization in the production of bioenergy and bioproducts from miscanthus. The OPTIMISC consortium consists of twelve partners in and outside Europe. The main objective of this project is to optimize miscanthus bioenergy and bioproduct chains by trialling elite germplasm types over a range of sites across Europe, Ukraine, Russia and China. The project is divided in eight work packages. The work of this thesis is related to work package 6, whose objectives are:

- Characterize the quality of biomass for the production of different biobased products, in various biorefinery chain.
- Evaluate the influence of different harvest regimes on the quality of miscanthus biomass.
- Assess the impact of abiotic stresses on the quality of biomass.
- Assess the influence of climatic conditions and geographic location on quality of Miscanthus biomass.

Fifteen elite genotypes of miscanthus (Table 1-1) have been established in six locations across Europe (Figure 1-1). Three replicates of each genotype were established across randomized plots in each
location. Samples from these trials are being analysed throughout three years. Samples from year 1 (Y1) have been analysed concerning the quality of the feedstock for bioethanol production to build prediction models for cell wall composition and saccharification efficiency based on near-infrared spectroscopy (NIRS) data (Section 1.6 of this thesis), a high-throughput technology that allows to generate reliable data in a fast way. In this thesis, samples collected in year 2 (Y2) were analysed for this same characteristics in order to expand the calibration model of Y1.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Genotype</th>
<th>Species</th>
<th>Propagation</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OPM-1</td>
<td><em>M. sacchariflorus</em></td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>2</td>
<td>OPM-2</td>
<td><em>M. sacchariflorus</em></td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>3</td>
<td>OPM-3</td>
<td><em>M. sacchariflorus</em></td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>4</td>
<td>OPM-4</td>
<td><em>M. sacchariflorus</em></td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>5</td>
<td>OPM-5</td>
<td>Hybrid</td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>6</td>
<td>OPM-6</td>
<td>Hybrid</td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>7</td>
<td>OPM-7</td>
<td>Hybrid</td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>8</td>
<td>OPM-8</td>
<td>Hybrid</td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>9</td>
<td>OPM-9</td>
<td><em>M. x giganteus</em></td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>10</td>
<td>OPM-10</td>
<td><em>M. sinensis</em></td>
<td>In vitro</td>
<td>JKI</td>
</tr>
<tr>
<td>11</td>
<td>OPM-11</td>
<td><em>M. sinensis</em> Goliath</td>
<td>In vitro</td>
<td>IBERS</td>
</tr>
<tr>
<td>12</td>
<td>OPM-12</td>
<td><em>M. sinensis</em></td>
<td>Seeds</td>
<td>IBERS</td>
</tr>
<tr>
<td>13</td>
<td>OPM-13</td>
<td><em>M. sinensis</em></td>
<td>Seeds</td>
<td>WU</td>
</tr>
<tr>
<td>14</td>
<td>OPM-14</td>
<td><em>M. sinensis</em></td>
<td>Seeds</td>
<td>WU</td>
</tr>
<tr>
<td>15</td>
<td>OPM-15</td>
<td><em>M. sinensis</em></td>
<td>Seeds</td>
<td>IBERS</td>
</tr>
</tbody>
</table>

Table 1-1. List of the fifteen genotypes established in the OPTIMISC multilocation trials.

Figure 1-1. Map indicating the locations of the OPTIMISC multi-location trials.
1.6 Near-infrared spectroscopy (NIRS)

Near infrared spectroscopy (NIRS) is a fast analytical method that can provide multi-constituent analysis of any matrix (Reich, 2005). A NIR spectrometer is usually composed by a light source, a monochromator, a sample holder and a detector. Solid samples can be measured by diffuse transmittance, diffuse reflectance or transflectance. The detector measures the absorbance (A) values of each sample.

NIRS analysis for cell wall composition is done using dry, grinded stem material (Haffner et al., 2013). The spectral data obtained by this technique is complex and requires the use of mathematical pre-treatments in order to be able to predict the composition characteristics of the samples. These pre-treatments are mathematical corrections made to reduce, eliminate or standardize the impact of factors like light scattering, path length variations and random noise (Reich, 2005).

To perform any quantitative analysis with a NIR spectrometer, it is necessary to first build a multivariate calibration model, which involves four steps: i) the selection of a representative calibration sample, ii) the spectra acquisition and determination of reference values, iii) the multivariate modelling to relate the spectral variations to the reference values, and iv) the validation of the model (Reich, 2005).

In order to build an efficient calibration model, two set of samples are needed (Figure 1-2). The size of such sets may vary according to the complexity of the spectral data and the component to be analysed. Firstly, a calibration set of samples is used to develop a calibration model, which is done by correlating the spectral data to the different parameters obtained by a chemical method. The correlation must be condensed to a minimum amount of factors that can describe the variation of the data points (Haffner et al., 2013). This is achieved by applying statistical tools such as the principal component analysis (PCA). After the calibration model has been built, it needs to be verified by using a validation set of samples or cross-validation. If this process yields a good prediction of the validation sample, or the cross-validation has good correlation with the results of the chemical method, the model can be used to predict the composition of the rest of the samples.

When using PCA on the spectral data, two statistical values are of special importance, namely global H (standardized Mahalanobis distance; GH) and neighbourhood H (NH) (Figure 1-3). The former refers to the distance of a sample from the centre of the spectral hypersphere. The latter indicates the closeness of a sample to the neighbouring samples (Hodgson et al., 2010) (Figure 1-3). These values help to identify outliers and are particularly useful to select samples for expanding prediction models. In the case of the study described in this thesis, these values were used to identify the samples from Y2 to build a calibration and a validation set to expand the model made in Y1.
Analysis of miscanthus biomass for biofuel production

Figure 1-2 (Xu et al., 2013). Flowchart for the development of prediction models to predict composition using NIR spectral data.

Figure 1-3. Graphic representation of the global H (standardized Mahalanobis distance; GH) and neighbourhood H (NH).
1.7 Thesis objectives

The general objective of this thesis project was to analyse the biomass quality for biofuel production of the fifteen elite genotypes established in the OPTIMISC multi-location trials for the second year. The effect of the location, year and genotypes on the saccharification efficiency was studied.

The specific objectives included:

- To obtain spectral data of the samples of the multi-location trials of the second year using near-infrared (NIR) spectroscopy.
- To use the prediction model built from the first year harvest to estimate the cell wall composition of the second year.
- To identify outliers in the samples, and determine the cell wall composition by biochemical methods.
- To expand the existing calibration model using the spectral data from the outliers.
- To validate the new calibration model.
2. Materials and methods

2.1 Plant material

Plant samples from five of the six locations of the OPTIMISC project corresponding to Y2 were used in this study. Material from the Potasch field (Ukraine) was not available at the time of this study. Plants were harvested after the winter, in 2014 and analysed at Wageningen University between October and December, 2014. Samples from each location consisted of stem sections of the fifteen elite genotypes, which were dried at 70°C and ground using a hammer mill with a 1mm sieve to obtain a fine powder. Table 1-1 shows the complete list of genotypes analysed.

2.2 Spectral data acquisition.

Powder samples were analysed using a NIRS DS2500 (FOSS, Denmark). Spectral data from a total of 223 samples was obtained. All data was labelled following the same format used in Y1. The name on each sample consisted of the letters “Opt” indicating the name of the project (OPTIMISC), followed by the three first letters of the location, followed by the number “2” indicating that the samples were taken on Y2, followed by the characters “plot” and finally, the plot number from which each sample was collected.

2.3 Sample selection for calibration and validation.

Analysis of the spectral data was carried out using the WinISI II Calibration Software (FOSS and Infrasoft International, Denmark). The prediction model built from Y1 data was used on the Y2 spectral data to identify outliers. To do this, the function “Compare spectra and equations” was applied using the .eqa file from Y1 and selecting the .nir file from Y2. In this way, samples with GH (see section 1.6) higher than three or NH higher than 1.7 were considered outliers and used as the calibration set of samples. A total of 37 outliers were detected. Additionally, a validation set of 25 samples was built by randomly selecting five samples from each location. The outlier samples and the validation set (a total of 62 samples) were analysed in the laboratory.

2.4 Dry matter content (DMC) and neutral detergent fibre (NDF) analysis

Between 490 and 500mg of powder from the selected samples were weighted inside ANKOM F57 filter bags (Ankom Technology, U.S.A.) using an analytical balance. Three replicates of each sample and eight control samples were weighted. Gravimetric analysis was used to determine the dry matter content (DMC) of the samples. The filter bags containing the samples were dried in an oven at 103°C overnight and weighted immediately after taken out. To calculate DMC, Formula 1 was used.
Analysis of miscanthus biomass for biofuel production

\[ DMC\% = \frac{Total\ dry\ weight - Bag\ weight - 0.001089}{Sample\ weight} \times 100\% \]

**Formula 1.**

Neutral detergent fibre (NDF) analysis was carried out using the filter bag technique and following the protocol indicated by the provider. To determine NDF as a percentage of DM (NDF%DM) formula 2 was used.

\[ NDF\%DM = \frac{Weight\ after\ NDF\ treatment - Bag\ weight + 0.000053}{Dry\ sample\ weight} \times 100\% \]

**Formula 2.**

Average values per sample (three replicates) were used for further data analysis.

2.5 Saccharification analysis

Saccharification characteristics after a mild alkaline pre-treatment were determined by measuring xylose and glucose release after 24 and 48 hours. As in the case of the NDF analysis, between 490 and 500 mg of powder were weighted inside ANKOM F57 filter bags using an analytical balance. Three replicates of each sample and nine controls were weighted.

Four steps were followed for this study namely wash of soluble sugars, alkaline pre-treatment, neutralization and enzymatic saccharification. Firstly, the washing step was done in an ANKOM 2000 Fiber Analyser (ANKOM, U.S.A.) using demiwater in 4 washing steps at 50°C. In the first two washes, α-amylase was added. After the wash, filter bags were put inside 50ml Eppendorf tubes together with 15ml of 2% NaOH and incubated for two hours at 50°C and 160rpm. A neutralization step followed by discarding the pre-treatment liquor and washing the samples twice with 40 and 45ml of demiwater respectively, for 5 minutes at 50°C and 160rpm. After discarding the washing liquor, 14ml of 0.1M sodium citrate buffer were put in the tubes and left in the incubator at the same settings for five minutes, after which the buffer was discarded. Finally, enzymatic saccharification was done by adding 44ml of 0.1 sodium citrate buffer, 0.3ml of Accellerase 1500 (Genencor, The Netherlands), and 15µl of Xylanase (Megazyme, Ireland) to each tube. Tubes were shaken and then incubated at 50°C and 160rpm. Samples for saccharification were taken 24 and 48 hours after the start of the incubation. Around 2ml of each sample were taken using a syringe and filtered using a 0.45µm filter and stored into a 2ml Eppendorf tube. To inactivate the enzymes, samples were heated to 95°C for five minutes.

For each sample, glucose and xylose concentration was measured. For glucose, a D-glucose dehydrogenase kit (Boehringer-Mannheim, Germany) was used. This kit is based on the measurement of the amount of NADPH formed in the reactions described in Figure 1-1 by means of

M.I. Santillán Martinez
its light absorbance at 340nm. A sample dilution of 1:10 was used for this measurement. Absorbance was determined using a BIORAD 680XR Microplate Reader (BIORAD, U.S.A.). For each sample, two technical replicates were measured.

\[
\text{Absorbance difference} = \text{absorbance difference of samples} - \text{average absorbance difference of blanks}
\]

**Formula 3.**

\[
glucose \text{ concentration} = \frac{(0.302)(180.16)}{(6.3)(1.0)(0.01)(1000)} \cdot (\text{Absorbance difference})
\]

**Formula 4.**

\[
xylose \text{ concentration} = \frac{(0.297)(150.01)}{(6300)(1.0)(0.01)} \cdot (\text{Absorbance difference})
\]

**Formula 5.**

Figure 2-1. Reactions used to determine D-glucose content.

Xylose concentration was measured using a D-xylene, xylan and arabinoxylan assay kit (Megazyme, Ireland). Measurement of xylose with this kit is based on the stoichiometric relation of the amount of NADH formed in the reactions described in Figure 2-2. NADH formed was measured by the increase in absorbance at 340nm using a BIORAD 680XR Microplate Reader (BIORAD, U.S.A). A sample dilution of 1:5 was used for this measurement. Two technical replicates per sample were measured.

\[
\alpha - D - \text{Xylose} \xrightarrow{\text{XMR}} \beta - D - \text{Xylose}
\]

\[
\beta - D - \text{Xylose} + NAD^+ \xrightarrow{\text{β-XDH}} D \text{ xylonic acid} + NADH + H^+
\]

**Figure 2-2. Reactions used to determine D-xylose content.**

2.6 Saccharification data analysis

For each sample, the absorbance difference was calculated according to Formula 3. Glucose concentration was calculated using the Formula 4. Xylose concentration was calculated using Formula 5. Total production of glucose and xylose were calculated using Formulas 6 and 7, respectively.
Analysis of miscanthus biomass for biofuel production

\[ \text{Total production of glucose} = (\text{Glucose concentration})(44.315)(10) \]

**Formula 6.**

\[ \text{Total production of xylose} = (\text{Xylose concentration})(44.315)(5) \]

**Formula 7.**

Values for cellulose as a percentage of NDF (Cell%NDF) and hemicellulose as a percentage of NDF (Hem%NDF) to determine glucose and xylose content, respectively, were taken from the predicted values of the Y2 dataset using the Y1 prediction model.

Saccharification efficiency was expressed in terms of percentage of glucose conversion as expressed in Formula 8.

\[ \text{Saccharification efficiency} = \frac{\text{total production of glucose}}{\text{cell wall glucose content}} \times 100\% \]

**Formula 8.**

### 2.7 Selection of the prediction model

Data from xylose and glucose conversion after 48 hours were used as calibration and validation data. Three prediction models were tested using the dataset of Y2: *i)* the prediction model from Y1 (built with 135 samples), *ii)* the prediction model of Y1 expanded with the calibration set of Y2 and validated with the Y2 validation set (built with 172 samples), and *iii)* a cross-validated model using all the samples from Y2 (built with 62 samples). A mathematical treatment 2,6,4,1 was used to develop the cross-validated prediction equation. This same treatment was used to develop the prediction model with data from Y1.
3. Results and discussion

3.1 Prediction of dry matter content and neutral detergent fibre

Percentage of dry matter content (%DMC) and neutral detergent fibre as a percentage of dry matter content (NDF%DM) were determined in the laboratory over the 62 samples. These values were compared with those generated by prediction on the spectral data using a model built with data from Y1. For both parameters, a high correlation (r) value was found (Table 3-1). In the case of %DMC, the correlation between the values predicted from the spectral data and the values obtained in the lab yielded an r value of 0.93, while in the case of NDF%DM, the r value was 0.94. These results show that spectral data from NIRS is useful to determine these two parameters in miscanthus.

In order to obtain the values for cellulose and hemicellulose content of the samples, the prediction model from Y1 was used. Unfortunately, the time scope for this study was not sufficient to analyse the calibration and validation set for these parameters in the laboratory. Chemical analysis for this parameter is necessary to validate the predictions for cellulose and hemicellulose content.

3.2 Determination of glucose conversion for calibration and validation sets.

Saccharification efficiency in terms of percentage of glucose conversion was measured in the laboratory samples (Figure 3-2). Among the 62 samples analysed, values between 17.08 (Wag2Plot29) and 44.15 (Abe2Plot23) were obtained. The lowest value corresponding to the genotype Miscanthus x giganteus and the highest corresponding to Miscanthus sinensis. The variation among these samples were also indicative of the highly variable values obtained using the prediction model for the entire set of Y2.
Figure 3-2. Glucose conversion measured in the laboratory in the validation and calibration set of samples.
3.3 NIRS calibration models for saccharification efficiency

The statistics for the prediction models tested are shown in Table 3-1. Unfortunately, none of the prediction models for xylose conversion yielded significant r values with the data obtained in the lab. In order to increase the accuracy of the model, a bigger number of samples to build the calibration model might be needed. For xylose conversion, the cross-validated prediction model built from data obtained in Y2 yielded a higher r value than the one of the other two models (0.56 vs 0.24 and 0.2). This can indicate the high variation existing between the samples across the two years made it impossible to predict the xylose conversion accurately with the model based on Y1.

However, with the prediction models for glucose conversion, a high r value was obtained between the data from the laboratory and the one obtained using the Y1 model (r=0.84), the expanded Y1 model (r=0.85, Figure 3-3) and the cross-validated model (r=0.9, Figure 3-4). A slight improvement in the model from Y1 was obtained after expanding it with the Y2 calibration samples and validating it with Y2 data. The cross-validated model made using the 62 samples from Y2 was used to predict the saccharification efficiency for Y2.

Several studies have focused on the use of NIRS to determine cell wall composition (Arnoult & Brancourt-Hulmel, 2014). However, the results obtained in this experiment confirm that the NIRS spectral data can also be useful to predict the glucose conversion of the feedstock, reducing the costs on the biochemical analysis and producing reliable data in a high throughput way.

Table 3-1. Statistics of calibration and cross-validation models. ‘Model’ column indicate the model described. ‘Nr of samples’ refers to the number of samples used to build the calibration model. Glucose and Xylose con 48 refer to the correlation value (r) for these traits between the laboratory and the predicted values. ‘SEP’ refers to the standard error of the prediction.

<table>
<thead>
<tr>
<th>Model</th>
<th>Nr of samples</th>
<th>Glucose Con 48</th>
<th>SEP</th>
<th>Xylose Con 48</th>
<th>SEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction model Y1</td>
<td>135</td>
<td>0.84</td>
<td>2.83</td>
<td>0.24 (N.S.)</td>
<td>2.35</td>
</tr>
<tr>
<td>Prediction model Y1+Y2</td>
<td>172</td>
<td>0.85</td>
<td>2.43</td>
<td>0.20 (N.S.)</td>
<td>1.99</td>
</tr>
<tr>
<td>Cross validation model Y2</td>
<td>62</td>
<td>0.9</td>
<td>1.91</td>
<td>0.56 (N.S.)</td>
<td>1.23</td>
</tr>
</tbody>
</table>
3.4 Prediction of saccharification efficiency

Using the cross-validated model, it was possible to predict the saccharification efficiency in terms of glucose conversion for the whole set of samples.
In Y2, the genotypes with the highest saccharification efficiency corresponded to those belonging to the species *M. sinensis* (Figure 3-5, Table 3-2). Specifically, the average values of glucose conversion of genotypes OPM15 and OPM13 were the highest across the five locations. The genotypes OPM11 (*M. sinensis* Goliath) and OPM12 (*M. sinensis*) also had high values of glucose conversion, particularly those located in Aberystwyth and Moscow. The lowest average values across locations corresponded to the species *M. × giganteus* and *M. sacchariflorus*. The genotype OPM9 (*M. × giganteus*) had the lowest value for saccharification efficiency in almost all locations. The genotype OPM1 (*M. sacchariflorus*) also exhibited remarkably low average values.

*M. sinensis* has been proven to have lower content of lignin than *M. × giganteus* (*Hodgson et al.*, 2010). Although the latter is one of the highest yielding genotypes (*Heaton et al.*, 2010), high biomass production is associated with high lignin content (*Arnoult & Brancourt-Hulmel*, 2014), which is likewise linked to poor saccharification efficiency, especially in mild pre-treatments (*Hideno et al.*, 2013). High digestibility is an important trait in plants for lignocellulosic feedstock. Highly degradable feedstock will allow the use of pre-treatments of lower intensity and, by so, will reduce costs in the production of second-generation biofuels. For this reason, these results are relevant for selecting high-quality genotypes for biofuel feedstock.

When compared to the saccharification efficiency of Y1, the results from Y2 present a lower overall average value (29.3 on Y2 vs 38.21 on Y1). The effect of year, which integrates both environmental conditions and age (maturity) of the crop, was statistically significant (P<0.001, Table 3-3) for glucose conversion. The effect of the age of the crop can contribute greatly to the composition of the biomass, especially in the first years. *Arnoult et al.* (*Arnoult et al.*) have shown that the prediction of traits related to biomass composition in miscanthus are less reliable when data of the first year of cultivation is used. In the present study, correlation values between the data from Y1 and Y2 was low (Figure 3-7). Although for the samples from Adana and Wageningen, it yielded values of 0.69 and 0.61, respectively, for the other five locations, the values remained below 0.5. Taking this into account, this results show that selection for saccharification efficiency may not be reliable in the first year, as the trait can vary greatly after this time.

It will be interesting to analyse the environmental conditions of locations such as Adana (Turkey), where the values for saccharification efficiency changed very few within the two years and compare them with those of locations like Wageningen, where such values were much lower in Y2. Furthermore, the plots corresponding to Adana (Turkey) were established before the rest of the locations. The maturity of the plants might have had a substantial impact on the stability of the trait along both years in this location.
Table 3.2: Average values for saccharification efficiency in terms of glucose conversion of plants belonging to the fifteen genotypes across five locations, throughout two years. Highest values indicated with green and lowest values indicated with red.

<table>
<thead>
<tr>
<th>Year</th>
<th>Location</th>
<th>OPM 1</th>
<th>OPM 2</th>
<th>OPM 3</th>
<th>OPM 4/5</th>
<th>Hybrid 1</th>
<th>Hybrid 2</th>
<th>Hybrid 3</th>
<th>sin 1</th>
<th>sin 2</th>
<th>sin 3</th>
<th>sin 4</th>
<th>sin 5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adana (TR)</td>
<td>22.51</td>
<td>28.18</td>
<td>24.55</td>
<td>30.09</td>
<td>29.14</td>
<td>51.08</td>
<td>26.74</td>
<td>29.38</td>
<td>29.99</td>
<td>30.11</td>
<td>34.38</td>
<td>33.68</td>
<td>31.18</td>
</tr>
<tr>
<td></td>
<td>Stuttgart (DE)</td>
<td>37.91</td>
<td>44.85</td>
<td>40.94</td>
<td>44.21</td>
<td>42.38</td>
<td>43.89</td>
<td>32.05</td>
<td>32.67</td>
<td>41.70</td>
<td>38.91</td>
<td>40.54</td>
<td>40.81</td>
<td>38.82</td>
</tr>
<tr>
<td></td>
<td>Wageningen (NL)</td>
<td>39.52</td>
<td>40.70</td>
<td>35.59</td>
<td>37.08</td>
<td>35.70</td>
<td>37.62</td>
<td>34.69</td>
<td>37.30</td>
<td>31.11</td>
<td>34.99</td>
<td>36.09</td>
<td>38.06</td>
<td>35.87</td>
</tr>
<tr>
<td></td>
<td>Aberystwyth (UK)</td>
<td>40.46</td>
<td>43.32</td>
<td>44.42</td>
<td>44.29</td>
<td>41.39</td>
<td>47.79</td>
<td>39.39</td>
<td>47.02</td>
<td>46.66</td>
<td>44.54</td>
<td>46.22</td>
<td>42.52</td>
<td>39.43</td>
</tr>
<tr>
<td></td>
<td>Moscow (RU)</td>
<td>38.17</td>
<td>44.23</td>
<td>38.87</td>
<td>41.46</td>
<td>41.00</td>
<td>43.83</td>
<td>40.86</td>
<td>45.03</td>
<td>39.88</td>
<td>39.91</td>
<td>43.06</td>
<td>44.50</td>
<td>41.82</td>
</tr>
<tr>
<td></td>
<td><strong>Average</strong></td>
<td>35.71</td>
<td>40.25</td>
<td>36.87</td>
<td>39.42</td>
<td>38.03</td>
<td>40.77</td>
<td>34.75</td>
<td>40.09</td>
<td>35.13</td>
<td>37.18</td>
<td>39.98</td>
<td>39.65</td>
<td>37.82</td>
</tr>
<tr>
<td>2</td>
<td>Adana (TR)</td>
<td>25.67</td>
<td>30.03</td>
<td>27.14</td>
<td>28.38</td>
<td>27.87</td>
<td>29.84</td>
<td>29.46</td>
<td>28.55</td>
<td>22.95</td>
<td>29.96</td>
<td>31.24</td>
<td>30.11</td>
<td>29.54</td>
</tr>
<tr>
<td></td>
<td>Stuttgart (DE)</td>
<td>24.50</td>
<td>29.57</td>
<td>21.96</td>
<td>27.76</td>
<td>27.26</td>
<td>28.98</td>
<td>25.49</td>
<td>25.81</td>
<td>22.98</td>
<td>26.50</td>
<td>28.79</td>
<td>29.54</td>
<td>29.71</td>
</tr>
<tr>
<td></td>
<td>Wageningen (NL)</td>
<td>24.36</td>
<td>26.55</td>
<td>24.60</td>
<td>24.48</td>
<td>23.76</td>
<td>23.72</td>
<td>23.02</td>
<td>23.66</td>
<td>17.95</td>
<td>24.64</td>
<td>23.74</td>
<td>24.71</td>
<td>27.48</td>
</tr>
<tr>
<td></td>
<td>Aberystwyth (UK)</td>
<td>31.19</td>
<td>34.94</td>
<td>32.40</td>
<td>33.33</td>
<td>33.03</td>
<td>31.88</td>
<td>32.23</td>
<td>32.20</td>
<td>29.97</td>
<td>28.56</td>
<td>39.04</td>
<td>34.71</td>
<td>36.38</td>
</tr>
<tr>
<td></td>
<td>Moscow (RU)</td>
<td>28.71</td>
<td>34.34</td>
<td>31.01</td>
<td>31.16</td>
<td>31.89</td>
<td>34.17</td>
<td>30.52</td>
<td>32.80</td>
<td>25.80</td>
<td>31.26</td>
<td>33.11</td>
<td>34.72</td>
<td>34.49</td>
</tr>
<tr>
<td></td>
<td><strong>Average</strong></td>
<td>26.84</td>
<td>31.09</td>
<td>27.42</td>
<td>29.02</td>
<td>28.76</td>
<td>29.72</td>
<td>28.14</td>
<td>28.59</td>
<td>23.03</td>
<td>28.34</td>
<td>31.19</td>
<td>30.76</td>
<td>31.88</td>
</tr>
</tbody>
</table>
It is also interesting to notice that in Y1 (Figure 3-6, Table 3-2), the highest values for saccharification efficiency were not among the *M. sinensis* genotypes, as in Y2. The difference between both years could have been caused by a combination of the maturity of the plant and the environmental conditions along both years. However, *M. x giganteus* was consistently, in both years, the genotype with the lowest average glucose conversion value across the locations. This data is of remarkable importance for plant breeding. Although *M. x giganteus* is a high-yielding plant, its glucose conversion percentage is lower than many other genotypes. The data obtained in this study is thus relevant for selection purposes. Breeding targets must be set in order to optimize both the yield capacity and the digestibility of the biomass of miscanthus in order to obtain varieties of high quality for biofuel production.

The effect of the location on glucose conversion was also proved to be significant for data from both years (P<0.001, Table 3-3). Indeed, it has been proven that the environmental conditions can significantly affect the cell wall composition of miscanthus (Hodgson *et al.*, 2010), which may ultimately affect its saccharification efficiency. In both years, plants established in Aberystwyth and Moscow yielded the highest glucose conversion values. The plants established in Wageningen were among the ones with the lowest glucose conversion value in both years. Interestingly, plants from Adana yielded the lowest glucose conversion percentage in Y1 but its average value changed very few in Y2. For this reason, it will be interesting to compare the environmental data, together with data regarding the quality of the soil in the different locations to understand the effect of these factors on the saccharification efficiency. A high variation among the replicates was also identified in some of the genotypes established in Aberystwyth, possibly due to a soil depth gradient.

Interactions between genotype, location and year were also proven to be significant for glucose conversion (P<0.001, Table 3-3). This further confirms the importance of selecting adequate genotypes that are able to grow optimally in specific conditions and locations.
Figure 3-5. NIRS predicted genotype means for glucose conversion for the second year.

Figure 3-6. NIRS predicted genotype means for glucose conversion for the first year
Analysis of miscanthus biomass for biofuel production

Figure 3-7. Correlation between the predicted values for glucose conversion of the first and second year by location.

Table 3-3. Analysis of variance of the effect of year, location and genotype on glucose conversion.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>(m.v.)</th>
<th>s.s.</th>
<th>m.s.</th>
<th>v.r.</th>
<th>F pr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep stratum</td>
<td>2</td>
<td></td>
<td>13.153</td>
<td>6.576</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>Rep.<em>Units</em> stratum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>1</td>
<td></td>
<td>8934.616</td>
<td>8934.616</td>
<td>2088.65</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Loc</td>
<td>4</td>
<td></td>
<td>5938.944</td>
<td>1484.736</td>
<td>347.09</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>OPM</td>
<td>14</td>
<td></td>
<td>1539.506</td>
<td>109.965</td>
<td>25.71</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Year.Loc</td>
<td>4</td>
<td></td>
<td>2485.806</td>
<td>621.452</td>
<td>145.28</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Year.OPM</td>
<td>14</td>
<td></td>
<td>324.25</td>
<td>23.161</td>
<td>5.41</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Loc.OPM</td>
<td>56</td>
<td></td>
<td>761.286</td>
<td>13.594</td>
<td>3.18</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Year.Loc.OPM</td>
<td>54</td>
<td>-2</td>
<td>477.235</td>
<td>8.838</td>
<td>2.07</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Residual</td>
<td>280</td>
<td>-18</td>
<td>1197.758</td>
<td>4.278</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>429</td>
<td>-20</td>
<td>20062.79</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Conclusions and future prospects

In this study, miscanthus plants from the OPTIMISC multi-location trials were analysed on the level of cell wall composition and saccharification efficiency. The results of this study provide relevant information on the variability of these traits on elite miscanthus genotypes. More importantly, these results will be useful to expand the NIRS models to predict cell wall composition and saccharification efficiency. The prediction model built with data from the first year for dry matter content and content of neutral detergent fibre yielded good correlation with the results obtained in the laboratory from the samples of the second year. This further confirms the efficiency of using NIR spectral data to determine cell wall composition.

In the case of saccharification efficiency, the prediction model for glucose conversion from the first year, the expansion of this model using data from the second year, and a new cross-validated model built only with the latter data set were all able to reliably predict the values for the second year. Unlike the model for xylose conversion, and despite the high variation between values of both years, the model for glucose conversion built from data of the first year was able to reliably predict the values for the second year. This has important implications on the methods used to measure saccharification efficiency, as it shows that NIR spectral data can be used not only to predict composition traits of the feedstock but also glucose conversion quality, proving the usefulness of this type of analysis to predict data related to one step forward on the bioethanol process chain.

Breeding miscanthus for biomass quality targeting the production of bioethanol is an ongoing effort. Most of the studies in this area are based on the yield and composition of the biomass. The results of this study further confirm the variability of saccharification efficiency across genotypes, locations and time. Breeding objectives must be set in order to maximise the digestibility of the feedstock without compromising the yield capacity. Fortunately, the genetic diversity of miscanthus species is vast and should be used to develop varieties that meet the requirements of the industry. Additionally, the standardization of the methods to determine saccharification efficiency must be set as a priority in order to optimize the research efforts towards the understanding of this trait.

The production of second generation biofuels still faces bottlenecks that need to be solved in order to achieve a commercial production. The results of this study confirm that miscanthus offers potential for breeding varieties whose use can improve the efficiency and lower the costs of the pre-treatments processes.
5. References


Arnoult S, Mansard M-C, Brancourt-Hulmel M, Péronne Cedex F. Early prediction of Miscanthus biomass production and composition based on the first six 1 years of cultivation 2.


Analysis of miscanthus biomass for biofuel production


Acknowledgements

I want to thank Luisa Trindade for giving me the chance to join the Bio-based Economy group and the OPTIMISC project and for her supervision during my work on this thesis. I am very thankful to Tim van der Weijde for his daily supervision, for giving me the motivation to carry out this work and for keeping every step of the process interesting and challenging. In this short period of time I was able to get involved and contribute in a topic that was completely new for me. Thanks to both for guiding me through this process.

I also want to thank all the people at the Bio-based Economy group. In these few months, I looked forward to the group meetings and presentations and I learned a lot from all of them. Thanks also to my colleagues in the Plant Breeding student room.

I thank the Executive Board of Wageningen University for granting me the WU Scholarship to course my MSc programme. I also thank the National Council for Science and Technology (CONACYT) and the Council for Science, Technology and Innovation of the State of Hidalgo (CITNOVA) for providing me the means to study this MSc programme.

This thesis is my final effort to obtain my Master degree. For that reason, I want to thank my dad, who taught me from a very young age to love plants. And my mom, who taught me that good answers can only lead to more and better questions.