Development of prediction models through Near-Infrared Spectroscopy (NIRS) and Biochemical Analysis for Hemp Cell Wall Content and Lignin

Development of prediction models through Near-Infrared Spectroscopy (NIRS) and Biochemical Analysis for Hemp Cell Wall content and lignin

Li Gu

MSc Plant Science-Plant Breeding and Genetic Resources

Wageningen UR

Thesis Supervisor: dr.ir. Luisa Trindade, MSc Jordi Petit Pedró (PhD candidate)

Register number: 911115287030 e-mail: li.gu@wur.nl

Date: February 2017

Abstract

Hemp (*Cannabis sativa* L.) as one of the fastest growing plants, and it was one of the first plants used for obtain fibres around 10,000 years ago. Hemp becomes a very important multi-purpose commercial crop now, which is used for paper, textiles, plastic, clothing, painting, biofuel, food and animal bedding. As hemp not only has high yield and produces high quality fibres, it is also an environmentally friendly crop with low requirements on fertilisers and virtually free from pests. High cellulose content, low component of lignification and reducing the cross links between pectin and the structural components of the cell wall are required for high quality fibre. To study the hemp cell wall composition and lignin, prediction models were developed by using NIRS spectrum data and biochemical data. The biochemical data were collected from Cell Wall Residue (CWR), Acid Detergent Lignin (ADL) and Klason Lignin (KL). After prediction models with good quality were developed, the prediction results showed that accessions, locations and their interactions had significant influence on cell wall content and lignin. The phenotypic variation had a stronger location influence than the genetic component on all the traits. However, the effect of the location is higher on the CWR% dm than on the lignin.

Keywords: Hemp, fibre quality, cell wall composition, lignin, NIRS, biochemical analysis, prediction model.

Table of Contents

1. Introduction

1.1 Overview of Hemp

Hemp (*Cannabis sativa* L.) has 2n=20 chromosomes (Mandolino et al., 1999), which belongs to *Cannabis* in Cannabaceae family. According to Hillig's research (2015), there are three main species of *Cannabis*: *Cannabis sativa, Cannabis indica* and *Cannabis ruderalis*. Hemp is considered to be originated in western and central Asia, including Russia, China, India, Pakistan and Iran (Anwar et al., 2006). As it is one of the fastest growing plants, and was one of the first plants used for obtain fibres around 10,000 years ago (Tourangeau, 2015), hemp becomes a very important commercial crop now, which is used for paper, textiles, plastic, clothing, painting, biofuel, food and animal bedding (Keller, 2013). Although the cultivation of hemp has reduced because of new competitors like cotton and synthetic fibres, production of hemp still reached almost 70,000 metric tonnes in 2013 (FAO).

Normally, hemp is planted between March and May in the northern hemisphere and between September and November in the southern hemisphere, and it needs 3-4 months to mature. Compared to other crops, hemp is an ideal fibre crop. It not only has high yield, good quality of fibre, and low requirements on fertiliser, but also it can fit well in crop rotation schemes, which contributes a lot on soil structure improvement (Du Bois, 1982; Hanson, 1980; Toonen et al., 2004). According to a 1998 study in *Environmental Economics,* hemp is considered as an environmentally friendly crop due to decrease of land use and other environmental impacts. Besides, growing hemp can help suppress weed growth and hemp is virtually free from pests so it can be grown without pesticides (Toonen et al., 2004; Van der Werf, 2004). However, hemp is vulnerable to various pathogens, such as bacteria, fungi, nematodes and viruses. These diseases do not reduce the yield of hemp but damage the fibre quality and stunted growth (T. Randall et al., 2004). Apart from diseases effects, low temperature, poor soil structure and unreasonable of water application are also problems for hemp plant establishment. Despite this, hemp is still a sustainable plant compared to cotton. Because the latter needs intensive use of pesticides, higher fertiliser and much stricter irrigation requirements, which causes many negative effects on the environment. China used to be the biggest hemp production country, however it was replaced by France recently. The production of hemp in France takes up more than 70 percent of the whole world output, followed by China with approximately 25 percent (FAO). According to FAO, an optimum yield of hemp fibre is more than 2 metric tonnes per ha, while the current yield average is only 650 kg/ha (Graeme, 2009).

1.2 Hemp Fibre

Natural fibres are roughly divided into two categories, woody and non-wood fibres (Van den Broeck et al., 2008). In wood, a fibre is a single cell, and its properties depend on what types of cells and its function in the tree, also tree species. Non-wood fibres are the collections of individual cells and they are classified depend on which part of plant they are found (Stevens et al., 2010).

The main contents of hemp fibre are cellulose, lignin and matrix polysaccharides, including hemicelluloses and pectin, which are associated with cellulose and lignin. However, the proportion of these main components depends on the fibre-type, climate conditions during vegetative period, cultivar, growing technology and plant parts. There are also some nonstructural components in hemp fibre, such as waxes, inorganic salts and nitrogenous substances (Dupeyre et al., 1998).

Hemp fibre is made up by two parts: bast and core (shives). Bast fibre belongs to non-wood fibres, located at inner bark, which is the outside part of the vascular cambium, and it is the strongest and stiffest fibre (Pickering et al., 2007). Bast fibres consist of many bundles of elementary fibres. The cell wall of the elementary fibre is made of a primary and a secondary layer, about 20-50mm long with a pericyclic form and thick cell walls from 5 to 15 µm (De Meijer, 1994; De Meijer and Keizer, 1994; Mediavilla et al., 2001; Toonen et al., 2004; Van der Werf et al., 1994b). In addition, secondary cell wall is the major proportion of bast fibres, which contains about 55% cellulose, 16% hemicelluloses, 8% pectin and 4% lignin (Hughes, 2012). The woody core which derived from the xylem tissue is about 0.5-0.6 mm long (De Meijer et al., 1994; Van der Werf et al., 1994), and it has significant higher content of lignin (~15%) (Hughes, 2012).

In the cell wall, cellulose is mainly formed of microfibrils which are highly ordered bundles of cellulose polymers and embedded in a matrix of other polysaccharides and lignin (Hughes, 2012). The microfibrils strongly helical in the secondary cell wall can be separated into three parts; S1, S2, S3. The S1 layer controls fibre stability by limiting excessive lateral cell expansion, meanwhile S3 layer resists hydrostatic pressure within the cell. As for S2, which is strongly influences the axial tensile properties of the fibre, the winding angle of it in hemp bast is normally lower than 10°, thus hemp bast has higher strength and stiffness. (Booker and Sell, 1998; Mark, 1967; Thygesen et al., 2007).

Figure 1.1 The structure of hemp cell wall (Hughes, 2012)

1.3 Hemp Application

Recently, natural vegetable fibres have been playing an increasingly important role in our day-to-day life. It has brought us a lot of benefits, such as economic viability, enhanced energy recovery, good biodegradability, low density (Dhakal et al., 2007; Le Troedec et al., 2008). However, natural fibres also have some disadvantages, like their physical and chemical properties are strongly dependent on genotype, harvest and environment (Le Troedec et al., 2008; Van de Weyenberg et al., 2006). Among the natural fibres, hemp is one of the strongest and stiffest available natural fibres, therefore it has great potential for applications in bio-composite materials (Pickering et al., 2007; Pickering et al., 2005). Hemp has become popular around the world since it is an environmentally friendly and multipurpose commercial crop with high quality fibre.

Hemp fibre quality can be determined by chemical composition, fineness, mechanical and sorption properties (Kostic et al., 2008). The proportion of chemical components influences fibre's structure, morphology and flexibility (Zofija et al., 2015). Fibre with high cellulose content provides strength and stability. Hemicellulose contributes little to stiffness and strength of fibre. The lignin in the section of amorphic cellulose can create mechanical incrustations which contribute to fibre lignification (Waśko and Mańkowski, 2004). The high content of lignin increase stiffness, makes fibre more breakable, and reduces its divisibility and spinnability (Nykter et al., 2008; Waśko and Mańkowski, 2004). Pectin presents in the middle lamella between all types of cells and hold the fibres together (Love et al., 1994), including bind the bast fibres and shives together. Because high fibre quality is also defined by good decortication features which means to separate bast fibres and shives easily (Easson and Molloy, 1996), so as for chemical composition, high fibre quality means high cellulose content, low component of lignification and reducing the cross links between pectin and the structural components of the cell wall (Mandolino and Carboni, 2004). Therefore, increasing the cellulose content and decreasing the hemicellulose, pectin and lignin content can improve fibre properties.

The bast fibres of hemp with more cellulose are normally used to make fabric, rope, paper and insulation materials. For shives, which have more lignin are used for mulch, animal bedding, construction materials, bio-composites for cars, etc. Hemp seeds can be used to produce oil used for oil-based paints or human consumption; Or they can be directly used to feed animals like bird. (Keller, 2013)

Figure 1.2 Hemp's constituent parts and their uses (Robinson and Schultes, 1996)

1.4 Research Objectives

This thesis is part of the project "Multipurpose Hemp for industrial bio-products and biomass (MultiHemp)", which is cooperated by 22 companies, research institutes and universities from different countries all over the world. The whole project is divided into several workpackages, and WU (Wageningen University) is responsible for workpackage 2; 'Genome-wide association mapping for hemp breeding'. The purpose of this project is to phenotype the cell walls composition of 124 different hemp accessions to identify the genetic basis underlying hemp fibre.

The main objective of this MSc thesis is to analyse the cell wall proportion and the lignin content among the whole sample set through prediction models. This objective can be achieved by the following steps:

-Phenotype the cell wall and lignin content of 124 hemp accessions by high-throughput methods (Near-InfraRed Spectroscopy (NIRS))

-Phenotype biochemically the cell wall and lignin content of a subset of samples to develop prediction models.

-Predict the phenotype of the whole sample set

-Analyse the prediction data and find out the influence of the environment on the fibre quality.

2. Materials and Methods

2.1 Materials

In this thesis, there are 124 different hemp accessions collected from 16 different countries including Europe, China and Canada. These hemp plants were cultivated in three different locations, Rovigo (Italy), Chèvrenolles, Neuville-sur-Sarthe (France) and Westerlee (The Netherlands) at 45°N 11°E, 48°N 0.2°E and 53°N 6°E. These 124 accessions not only contain breeders' materials but also wild accessions. The aim of this set is to cover different traits, including morphological and quality. Each location has three randomly designed blocks. The experimental unit was 1 m² for each plot (1.5 m² for French field trial). Hemp accessions were harvested at full flowering and five stems were collected randomly from each plot for genotype.

Each sample was cut into top, middle and bottom parts at different heights, which has been done by other workpackage groups. They were chopped to 2cm and put into the oven for 1.5 hours at 60 °C. The dried materials were grinded to 1mm diameter by the grinder machine 'Pepping 200AN' and packaged in a zipper plastic bags separately.

2.2 Methods

2.2.1 Near Infrared Spectroscopy (NIRS)

The Near Infrared spectroscopy (NIRS) is a faster and non-destructive method (Xu et al., 2013) compared to traditional wet chemical once for biomass composition analysis, so NIRS provides more accurate results. Besides, it is also a low-cost analysis. Thus, it has become a general method widely used to detect plant constituents. NIRS analysis is based on vibrational spectroscopy that monitors changes in molecular vibrations intimately associated with changes in molecular structure (Reich, 2005). After absorbing radiation, oscillating dipole moment which associated with vibrating bonds will interact with the radiation, cause a change of the dipole moment (Xu et al., 2013). The NIR spectrum is the total absorption of many chemical bonds (Toonen et al., 2004). NIRS can provide information on molecular overtones and combinations of vibrations, because the most prominent absorption bonds occurring in the NIR spectrum are related to those (Xu et al., 2013). The prediction of hemp cell wall composition is based on a set of fully characterised samples which covered broad wavelengths in the spectrum (Toonen et al., 2004).

Figure 2.1 NIR schematic diagram (Xu et al., 2013)

In this thesis, instead of 1116 (124*3*3) samples, only 1034 grinded hemp samples were scanned by NIRSTM DS2500, because the rest samples were missing. After scanning, next step should be analysing biochemical data for all the samples which would be very expensive. Therefore, we need to pick out some representative samples to calibration and develop the prediction model. To do this, all the hemp spectrum data were extracted by using the software 'Mosaic Solo (NIRS DS2000)' and ranked base on their H distance by using the software 'WinISI Project Manager'. There are two standardized H distances, one is the Global H (GH), which shows how different of one sample is from the average of all samples (Shenk and Westerhaus, 1991). The spectrum with GH values higher than 3.0 was considered as an outlier. The other one is the minimum standardized H distance - the Neighbourhood H (NH). This value is the distance to the closest neighbouring sample, and it is used to control the closeness of neighbouring samples within the dataset (Olinger et al., 2001). If the NH distance from one sample to its nearest neighbour was higher than 1.5, then it also was considered as an outlier. At the end, 116 samples were selected from all the scanned hemp samples based on the variation of the NIR spectrum and were used for further biochemical analysis. Among these 116 hemp samples, 99 samples were model samples and 17 were outliers. What else, 25 samples were randomly selected from the rest 918 samples as validation samples.

2.2.2 Biochemical Analysis

The 141 selected samples were biochemically analysed by three different methods, analysis of the cell wall content using Alcohol Insoluble Solids (AIS) extraction method, analysis of the Acidic Detergent Lignin (ADL) using Goering and Van Soest's protocol (1970) (ANKOM technology), and analysis of the Acid Insoluble Lignin (Klason lignin) by using two-step hydrolysis procedure. The main difference between AIS and ANKOM technology is pectin

been included or not. For AIS, which uses 80% alcohol, pectin will remain while it is not soluble in alcohol (Filomena et al., 2012). So AIS is used for major cell wall compositions detection. In general, the ANKOM technology is used to determine the neutral detergent fibre (NDF), acid detergent fibre (ADF) and acid detergent lignin (ADL). As part of the pectin is soluble in water, %NDF is the sum of the cellulose, hemicellulose and lignin parts of the fibres by using a neutral detergent. %ADF and ADL%_dm are finished in one procedure using an acid detergent instead of a natural detergent. %ADF includes lignin and cellulose parts of fibres, whereas ADL% dm only has the lignin part of the acid detergent fibres (Toonen et al., 2004). %ADF procedure was done before and in this thesis, we mainly studied lignin by using ANKOM technology, so only ADL% dm was included. Both Klason Lignin and ADL are lignin, Klason is the standard method for wood analysis and ADL is the most common method for animal science and agronomy. Klason lignin was used two-step sulfuric acid hydrolysis while ADL was used sequential detergent analysis.

A. Alcohol Insoluble Solids (CWR)

The 141 samples were randomly divided into 7 batches, each batch had two biomass controls and one blank. This experiment has been done in two replicates. For each selected hemp sample, after proper mixed, $50(\pm 1)$ mg biomass has been weighted. The biomasses were put into 2ml microcentrifuge tube and labelled. AIS can be divided into two parts, cell wall extraction and Alpha-amylase digestion.

Cell wall extraction

First step, each tube added 1.5ml 80% ethanol and was put on ice for 30'. Every 15' the tubes should be mixed properly to make sure this procedure fully reaction. Then tubes were centrifuged for 5' at 10,000g and discard supernatant. Repeated the first step for three times. This step means to remove all other things including DNA, membrane, etc., but only left cell walls for next step. Second step is 1.5ml acetone was added in each tube and waited for 10'. After that, tubes were centrifuged for 5' at 10,000g and discarded supernatant. The acetone step is same for the methanol. We used acetone and methanol to remove chlorophyll and lipids. At the end, we put the tubes in the Thermomixer overnight to remove the methanol which would inactivate the alpha-amylase in the next part.

Alpha-amylase digestion

Before start with the digestion, Tris-maleate buffer needed to be unfrozen and mixed properly. Also, the amount of α -amylase enzyme was calculated and weighted. 300 μ l of 10Mm Tris-maleate buffer was added into each sample. Samples were put in the Thermomixer at 30°C (room temperature) for 30'. Then samples were placed in boiling water for 5'. Next, the samples needed to be equilibrate to 40°C because 40°C is the optimal temperature for α -amylase enzyme, so we put samples on the mixture of ice and water to cooldown. After that, 100µl of α-amylase enzyme solution with 2U/mg of carbohydrate was added to each sample and waited at 40°C (Thermomixer) for 1 hour. Afterward, another half dose of α-amylase enzyme solution was applied fir 30' at 40°C (Thermomixer). We added four volumes of cold absolute ethanol, which was approximately 1800µl, and precipitate polysaccharides at -20°C (Freezer) for at least 1 hour. Then samples were centrifuged at

2,000g for 5' and the supernatant was discarded. Repeated the cold absolute ethanol for three more times but just for wash. Finally, the samples we put in the Thermomixer overnight at 60°C to get rid of the ethanol, then the samples were dry in the oven at 103°C overnight. The dry weight of each sample was measured after oven.

B. Acid Detergent Lignin (ADL)

As the ADF was performed before this experiment, the ADF samples were directly used to perform the ADL. For this phenotyping 61 samples where completely analysed before but for the 80 resting samples, the ADF was only performed. In this thesis, the 80 samples were analysed in triplicates, so in total there were 240 ADF samples that were randomly splitted into 11 batches. Each batch included 22 samples, 1 control and 1 blank. We put one batch in a bottle with 500ml of 72% sulfuric acid. for 3 hours constantly shaking at room temperature. Samples were cleaned by demi water for three times until the pH was neutral. Then the washed samples were put in acetone for 5' and placed in the fume hood to dry. After at least two hours in the fume hood, the samples were transferred into the oven at 103°C overnight and then the dry weight of each sample was measured.

C. Acid Insoluble Lignin/Klason Lignin (KL)

The Klason lignin was analysed by using a two-step hydrolysis with sulfuric acid. The first hydrolysis step was adding 72% sulfuric acid to samples and put in the incubator for 1 hour at 30°C. The second hydrolysis step was diluted 72% sulfuric acid to 4% with MQ water and put into autoclave for 1 hour at 121°C. Then the samples were cooling at room temperature. Afterwards, the insoluble acid fraction of the sample was filtered with a glass filter (47 mm hydrophilic glass fibre filter with a 0.7 μ m pore size) and washed with demi water for several times. The residue on the glass filters was KL. The glass filters with residues were put into the oven for overnight at 103°C and dry weight of each sample was measured after oven.

D. Data analysis

The dry matter content data was provided, and the weight of empty filter bags/tubes, filter bags/tubes + samples and dry weight of filter bags + samples were measured. The controls were used to checked if each batch was correctly performed. Then all the data were corrected by the differences between initial weight and dry weight of blanks per trait. By taking into account of all the values, the percentage of CWR, ADL and KL were calculated.

Blank correction = Blank dry weight after oven − Blank weight before oven

 $DMC\% = \frac{Total\, weight - Filter\, bag\, weight - Blank\, correction}{\sqrt{O(1-\frac{1}{2})}}$ $\frac{1}{5}$ $\frac{1}{2}$ $\frac{1}{2}$

$$
\text{CWR\%_{_}dm} \text{ or } \text{ADL\%_{_}dm} = \frac{ (Filter \ or \ tube + Sample) dry \ weight - Filter \ or \ tube \ weight - Blank \ correction}{ Sample \ weight * DMC\%} * 100\%
$$

The standard deviation (SD) and the relative standard deviation (RSD) of all samples data for each replicate per trait were also calculated. All the samples with RSD higher than 15 were considered not reliable except for the KL%_dm, because the sample weight of KL was very low thus KL treatment easily caused big errors. Since the controls from KL were similar across batches, we considered that the data was reliable. For each trait, the data was

corrected in three different ways: data corrected by batch, data corrected by control and double correction data, plus the original data Then, these data were separated into two groups by sample type: calibration (models + outliers) and validation.

2.2.3 Prediction Models Development

Three prediction models (CWR, ADL and KL) were developed for each set of data (Original data, data corrected by batch, data corrected by control, data corrected by batch and control). The prediction models were developed by correlating the biochemical data and the NIR spectrum data per trait.

The modelling was performed by using the software 'WinISI Project Manager' for each trait. First, a Principal component analysis (PCA) was performed to find outliers from the calibration samples. In case that there were outliers, they would be purged from the calibration sample set. 7 components were used to measure GH which included 98.34% of their spectral variability. During this analysis, the outliers of the calibration samples should have been selected and purged. After this the Modified PLS (Partial Least Squares) regression method was performed to develop the equations. Then, the calibration samples were used to develop equations with the following settings of wavelengths and math treatment: SNV and Detrend (standard one), derivative '1', gap '4', smooth '4' and smooth 2 '1', and H or R measurement, H or R value was '3'. Then the 25 randomly selected validation samples were used for validate the model. To check the quality of the model, the R^2 (squared correlation coefficient) between the predicted values with the reference value were used. When the value of R^2 is close to '1', it means this is a good model. Also, the standard error of predicted data (SEP) and the standard error of lab data (SEL) were checked. When the value of SEP is lower than three times of SEL value (SEP<3*SEL), then it considered to be a good model. Finally, the best models from the original data, corrected data by batch, corrected data by control and double correction data were chosen and all 1034 samples data were predicted for each trait (CWR, ADL, KL).

2.2.4 ANOVA Analysis

Each genotype had 9 replicates (3locations*3biological replicates). We complemented the missing data with * and the data was analysed with the software 'GenStat'. Analysis of the variances were performed (General ANOVA, GxE), using the mean plot values to evaluate the components of the phenotypic variation, including variation attributed to genotype (G), location (E) and genotype x environment interactions (GxE).

3. Results

3.1 Biochemical analysis

As it can be seen from the Table 3.1, the T-test was done for CWR%_dm, ADL%_dm and KL% dm. There were 140 samples of CWR and KL, 141 samples of ADL. The average dry matter content value of CWR% was the highest which reached 89.57%, only 9.15% and 14.90% for ADL% dm and KL% dm separately. The variation of CWR% dm was from 81.39% to 95.38%. As for lignin, the variation of KL%_dm was larger than ADL%_dm, which was from 9.82% to 20.53% and 6.55% to 14.13% respectively. Furthermore, the phenotypic variation between samples was significantly different (P=0.000<0.001) for each trait.

Table 3.1. T-test for CWR%_dm, ADL%_dm and KL%_dm

As the sample variation of each trait was significant, thus all the samples were effectively used for prediction models development.

3.2 Prediction Models

Instead of 1116 samples, only 1034 hemp samples were available for model development of CWR%_dm, ADL%_dm and KL%_dm. The quality of the prediction models was shown as RSQ. The value of RSQ closer to 1, which means the model prediction value closer to real lab data, then the prediction model has a better quality. As the results of prediction models by using calibration samples showed in Table 3.2, 3.3 and 3.4, for CWR%_dm, the highest RSQ value was from the data corrected by batch, which reached 0.9014. For the original data, the RSQ of CWR% dm was 0.8788. And for ADL% dm model, the highest RSQ was 0.904 by using original data. As for KL% dm, the RSQs calculated by using original data and the data corrected by batch were close, which was 0.869 and 0.879 respectively.

Table 3.2 Developed models for CWR%_dm by using calibration samples

Because the RSQs of different correction data sets were close, so after the prediction models was developed, the validation samples were used to validate the models (and the validation results were showed in Table 3.5). The quality of the prediction models by using validation samples was shown as R^2 , which is the same as RSQ, just to distinguish the samples were used. The R^2 of CWR%_dm by original data model was 0.71, and the model based on the data corrected by batch had the highest R^2 , which was 0.713 close to 0.71. The R^2 of KL%_dm was lower than the R^2 of CWR%_dm. The highest R^2 of KL%_dm was 0.596 from the data corrected by batch. As for the original data, R^2 was 0.591. Among the three traits, ADL%_dm had the highest quality model that the R^2 of ADL%_dm were around 0.9, and between the four data-sets, the original data had the highest value of R^2 which reached 0.914. Even though CWR% dm and KL% dm had lower quality models, but compared to the previous prediction models, they had a higher value this time. So, there were some good quality prediction models for each trait which means these models would give high accuracy trait prediction values.

Table 3.5 Prediction models validated by validation samples.

A higher quality prediction model for each trait was needed, meanwhile the quality value of original data and the data corrected on batch were higher than the other two data sets. Finally, the prediction model of original data for each trait was chosen, and each prediction model had a relatively high quality.

Then the original biochemical data and predicted data were put into a figure for each trait to check the model quality which was more intuitive.

Figure 3.1 Original biochemical data vs. predicted data of validation samples of CWR%_dm. percent of CWR%_dm original data can be explained by the prediction model.

Figure 3.2 Original biochemical data vs. predicted data of validation samples of ADL%_dm. 90 percent of ADL%_dm original data can be explained by the prediction model.

Figure 3.3 Original biochemical data vs. predicted data of validation samples of KL%_dm. 59 percent of KL% dm original data can be explained by the prediction model.

3.3 ANOVA analysis

All the analysis of the variance was performed by using the predicted data from the original data set as the quality was good enough. The influences of genotype (accessions) and environment (locations) on the experiment data and if they had interactions were checked by general ANOVA model.

If the F value smaller than 0.001, then the influence of that treatment was significant. From the Table 3.6, 3.7 and 3.8, it can be observed that the F value of entry (accessions), locations and the interaction between them was smaller than 0.001 for CWR% dm, ADL% dm and KL%_dm. Thus, the influence of accessions, locations and the interaction between accessions and locations were significant for each trait.

Furthermore, from v.r. showed in the Table 3.6, 3.7 and 3.8, if the variance ratio value was larger, which means the influence was stronger. For CWR%_dm, the variance ratio of location was 1959.85 while the variance ratio of accessions was only 3.96. Compared to CWR%_dm, the variance ratio of location of ADL%_dm and KL%_dm were lower, which was 305.35 and 820.04 respectively. However, the ADL%_dm and KL%_dm variance ratio of accessions were 16.6 and 29.57, still lower than the variance ratio of locations. Thus, the influence of the locations was not only stronger than the influence of accessions and the interaction between accessions and locations, but it was also more significant on CWR%_dm than on ADL% dm and KL% dm.

Source of						
variation	d.f.	m.v.)	S.S.	m.s.	v.r.	F pr.
Entry	123		844.952	6.87		3.96×001
Location			6791.12	3395.56	1959.85 ≤ 0.001	
Entry.Location	227	-19	791.23	3.486	2.01	≤ 0.001

Table 3.6 ANOVA analysis of CWR%_dm

Table 3.7 ANOVA analysis of ADL%_dm

Source of						
variation	d.f.	(m.v.)	S.S.	m.s.	v.r.	F pr.
Entry	123		613.7785	4.9901		16.6 ≤ 0.001
Location			183.6204	91.8102	305.35×0.001	
Entry.Location	227		-19 133.6579	0.5888		1.96×001

Table 3.8 ANOVA analysis of KL%_dm

To check the influence differences of the environment between three locations (the Netherlands, France, and Italy) for CWR% dm, ADL% dm and KL% dm, all the prediction data were showed on a graph for each trait. For CWR%_dm, three locations had significantly difference. The data of the samples from the Netherlands had the highest value, followed by Italy, then France. But for KL%_dm and ADL%_dm, the data of the samples from Italy and France were similar, whereas the samples from the Netherlands still had the highest value.

Figure 3.4 Means for entry at different levels of location of CWR%_dm

Figure 3.5 Means for entry at different levels of location of ADL%_dm

Figure 3.6 Means for entry at different levels of location of KL%_dm

Due to the influence of three locations were different between the cell wall residue and the lignin, the correlations between CWR%_dm, KL%_dm, ADL%_dm were also calculated. The correlation values were shown in Table 3.9 and Table 3.10. It can be seen that CWR%_dm, KL% dm and ADL% dm were significantly correlated. ADL% dm and KL% dm were strongly correlated with a r^2 of 0.7507. KL%_dm and ADL%_dm were also correlated with CWR%_dm but only with a r^2 of 0.2701 and 0.3237 respectively.

		$CWR\%$ dm ADL% dm KL% dm	
CWR% dm			
ADL% dm	0.3237		
KL% dm	0.2701	0.7507	

Table 3.10 Two-sided test of correlations different from zero

4. Discussion

From biochemical analysis, it can be found out that the variation between the experiment samples was significantly different for each trait. Therefore, it was meaningful to make the prediction models as the second step.

As for prediction models, after RSQ of the calibration samples, R^2 of the validation samples were checked, all the prediction models were proved with good quality. The original data prediction model and other three corrected data prediction models had very similar evaluation values. The reason that we chose original data prediction model as our final prediction model because the original data model is more convenient to use in the further steps which the data do not need correction. Also, the original prediction model is much closer to the real experimental data. However, the original data model may have slightly discrepancy with ideal model.

From the ANOVA analysis result, it proved that genotype (accessions) and environment (locations) had significant influence on data. Also, there was interaction between genotype and environment. The ADL% dm and KL% dm data of the samples from the Netherlands was much higher than the samples from France and Italy. The CWR% dm value of the samples from the Netherlands was the highest, then followed by samples from Italy, the lowest was from France. Also, the analysed results showed that the same accessions cultivated in different places had different CWR%_dm, ADL%_dm and KL%_dm. Almost all the samples had the highest value when grown in the Netherlands, which might be explained by the effects of environment, especially the temperature and the light intensity, among these three locations on the biomass quality (Tang et al., 2016).

Hemp is sensitive to the temperature and the light intensity (Pahkala et al., 2008), and the temperature in the Netherlands is relatively lower than the temperature in France and Italy. As well as light intensity, in the Netherlands the light intensity is quite weak compared to France and Italy. Because of environment differences, the flowering time of hemp in the Netherlands is delayed because of lower temperature and weaker light intensity. The stem growth of the hemp will slow down after flowering (Van der Werf et al., 1994a), which means the harvesting time of hemp, when the maximum yield fibre reached, was delayed according to the flowering time (Amaducci et al., 2002; Mediavilla et al., 2001). Because the samples grown in the Netherlands had a delayed flowering time and a prolonged vegetative growth period, they produced higher yield of fibre (Van der Werf et al., 1994a). Thus, the CWR%_dm values of hemp from the Netherlands were the highest. As a part of cell wall, lignin (KL and ADL) were also produced more when cell wall content increased.

Since the influence of environment was more significant on CWR than lignin, the correlations between CWR% dm, ADL% dm and KL% dm were calculated. Among them, ADL% dm and KL% dm were strongly correlated. ADL% dm and KL% dm are all lignin means from different methods, but the value of KL% dm was higher than ADL% dm. KL molecular composition is similar with ADL, but the amount of ADL%_dm was underestimates due to loss of acid-soluble lignin in the acid detergent step of the procedure and KL could be overestimation (Kondo et al., 1987; Lowry et al., 1994). However, there was only slightly correlation between CWR% dm and lignin (ADL% dm and KL% dm), which means that with

the increase of CWR%_dm, the lignin content may also be increased, but mainly caused by other content of cell wall increased, such as cellulose, hemicellulose or pectin.

In conclusion, the plants that were cultivated in a lower temperature and weaker light intensity gave a higher amount of fibres; the plants that cultivated in a higher temperature and stronger light intensity showed lower amount of lignin. Therefore, they might increase the fibre quality as the stiffness of the fibre reduces with the drop of lignin.

However, this thesis had some limits. As we can see from biochemical analysis, CWR had lower RSD which means it had less experimental errors than the other two experiments. The reason why ADL and KL had higher RSDs and even some of their RSDs were reached more than 30 (which should be lower than 15) is possibly that each sample weight was quite low. Because lignin only takes a very small percentage in the cell wall composition, so ADL%_dm and KL% dm values were relatively small. Thus, even only a small difference may cause big errors on the results. So, to make less or avoid experimental errors, we should weigh enough sample weight. Also, to make sure about the data accuracy, at least three replicates should be included. If the data goes wrong, we need to repeat the experiment again. During the experiment, the procedure must keep the same.

At the end, the final purpose of this project is to determine the hemp cell wall composition and improve the fibre quality. In this thesis, we only analysed the general hemp cell wall residue and lignin, we still need to find out the percentage of cellulose, hemi-cellulose, pectin and their influencing factors with doing more kinds of experiments. At last, we can base on our purpose to choose the hemp we want and where should it grow, like high quality fibre with high cellulose and low lignin. So, there is still a long way to go.

5. Conclusion

In this thesis, 1034 samples of different accessions from three different locations were already scanned by NIRS to help developing the prediction model. Based on the NIRS, 140 samples were selected for CWR, 141 samples for ADL and 140 samples for KL. After finished biochemical analysis, each trait with experimental data was corrected. Biochemical data had large variation of CWR% dm, ADL% dm and KL% dm. Then the biochemical analysed data and NIRS spectrum were integrated together, the prediction model for each trait was developed separately. Also, the validation samples were used to validate the prediction model to exam the models quality. Prediction models with good quality were developed. At last, all the prediction data were put into ANOVA to find out the influencing factors and the correlations between CWR, ADL and KL. The ANOVA analysis indicated that the hemp grown in the Netherlands had the highest CRW% dm, ADL% dm and KL% dm, and the samples from Italy had higher CRW% dm than the samples from France. But for lignin, no matter the samples were from France or Italy, they had similar data. Also, lignin was strongly correlated with each other but not with CWR% dm. In conclusion, CWR% dm, ADL% dm and KL% dm was strongly affected by environment, accessions and their interactions. The phenotypic variation had a stronger location influence than the genetic component on all the traits. However, the effect of the location is higher on the CWR%_dm than on the lignin.

Acknowledgement

I would like to give my great thanks to my supervisors Luisa Trindade and Jordi Petit Pedró for giving me this chance to do this interesting hemp thesis project, also grateful for their patient guidance and helpful feedbacks during the whole thesis period. I would also like to thank to Heelen Furrer-Verhorst, for giving me help in the bio-chemical lab. At last, I want to thank all the people in the Bio-based Economy Group, it is my pleasure and honour to work and study with these people, I spent a great and enjoyable time in the last 6 months.

References

1. Amaducci, S., Errani, M., and Venturi, G. (2002). Plant population effects on fibre hemp morphology and production. *Journal of Industrial hemp 7*, 33-60.

2. Anwar, F., Latif, S., and Ashraf, M. (2006). Analytical characterization of hemp (Cannabis sativa) seed oil from different agro-ecological zones of Pakistan. *Journal of the American Oil Chemists' Society 83*, 323-329.

3. Booker, R., and Sell, J. (1998). The nanostructure of the cell wall of softwoods and its functions in a living tree. *Holz als Roh-und Werkstoff 56*, 1-8.

4. De Meijer, E. (1994). Variation of Cannabis with reference to stem quality for paper pulp production*. Industrial Crops and Products 3*, 201-211.

5. De Meijer, E., and Keizer, L. (1994). Variation of Cannabis for phenological development and stem elongation in relation to stem production. *Field Crops Research 38*, 37-46.

6. Dhakal, H., Zhang, Z., and Richardson, M. (2007). Effect of water absorption on the mechanical properties of hemp fibre reinforced unsaturated polyester composites. *Composites Science and Technology 67*, 1674-1683.

7. Du Bois, W. (1982). Hennep als grondstof voor de papierindustrie. *Bedrijfsontwikkeling.*

8. Dupeyre, D., and Vignon, M. (1998). Fibres from semi-retted hemp bundles by steam explosion treatment. *Biomass and Bioenergy 14*, 251-260.

9. Easson, D., and Molloy, R. (1996). Retting--a key process in the production of high value fibre from flax. *Outlook on agriculture.*

10. Filomena, A P., Cherie, W., Geoffrey, B., and Antony B. (2012). Determining the polysaccharide composition of plant cell walls. doi:10.1038/nprot.2012.081

11. Goering, H.K., and Van Soest, P.J. (1970). Forage fiber analyses (apparatus, reagents, prcedures, and some applications). *USDA Agr Handb*.

12. Graeme Thomas. "Natural Fibers: Hemp Food and Agriculture Organization of the United Nations, 2009".

13. Hanson, J. (1980). An outline for a UK hemp strategy. *The Ecologist 10*, 260-263.

14. Hillig, K.W., 2005. Genetic evidence for speciation in Cannabis (Cannabaceae). Genet.Resour. *Crop Evol. 52*, 161–180.

15. Hughes, M. (2012). Defects in natural fibres: their origin, characteristics and implications for natural fibre-reinforced composites. *Journal of materials science 47*, 599-609.

16. "Industrial hemp's double dividend: a study for the USA". *Ecological Economics*. 25: 291– 301. 1998-06-30. doi:10.1016/S0921-8009(97)00040-2. Retrieved 2013-11-09.

17. Keller, NM (2013), "The Legalization of Industrial Hemp and What it Could Mean for Indiana's Biofuel Industry", *Indiana International & Comparative Law Review*, 23 (3): 555, doi:10.18060/17887

18. K.Tang., P.C. Struik., X. Yin., C. Thouminot., M. Bjelková., V. Stramkale., S. Amaducci. (2016). Comparing hemp (Cannabis sativa L.) cultivars for dual-purposeproduction under contrasting environments. *Industrial Crops and Products 87 (2016) 33–44.*

19. Kondo, T., Minzuno, K., Kato, T. (1989). Variation in solubilities of lignin in acid detergent and in alkali. *J. Jpn. Grassl. Sci.* 33, 296-299.

20. Kostic, M., Pejic, B., and Skundric, P. (2008). Quality of chemically modified hemp fibers. *Bioresource Technology 99*, 94-99.

21. Le Troedec, M., Sedan, D., Peyratout, C., Bonnet, J.P., Smith, A., Guinebretiere, R., Gloaguen, V., and Krausz, P. (2008). Influence of various chemical treatments on the composition and structure of hemp fibres. Composites Part A: *Applied Science and Manufacturing 39*, 514-522.

22. Love, G., Snape, C., Jarvis, M., and Morrison, I. (1994). Determination of phenolic structures in flax fibre by solid-state 13 C NMR. *Phytochemistry 35*, 489-491.

23. Lowry, J. B., Conlan, L. L., Schlink, A. C., McSweeney, C. S. (1994). Acid detergent dispersible lignin in tropical grasses. *J. Sci. Food Agric. 65*, 41-49

24. Mandolino, G., and Carboni, A. (2004). Potential of marker-assisted selection in hemp genetic improvement. *Euphytica 140*, 107-120.

25. Mandolino, G., Carboni, A., Forapani, S., Faeti, V., and Ranalli, P. (1999). Identification of DNA markers linked to the male sex in dioecious hemp (Cannabis sativa L.). *Theoretical and applied genetics 98*, 86-92.

26. Mark, R.E. (1967). Cell wall mechanics of tracheids. Cell wall mechanics of tracheids.

27. Mediavilla, V., Leupin, M., and Keller, A. (2001). Influence of the growth stage of industrial hemp on the yield formation in relation to certain fibre quality traits. Industrial *Crops and Products 13*, 49-56.

28. Nykter, M., Kymäläinen, H.-R., Thomsen, A.B., Lilholt, H., Koponen, H., Sjöberg, A.-M., and Thygesen, A. (2008). Effects of thermal and enzymatic treatments and harvesting time on the microbial quality and chemical composition of fibre hemp (Cannabis sativa L.). Biomass and bioenergy 32, 392-399.

29. Olinger, J., Griffiths, P.R., and Burger, T. (2001). Theory of diffuse reflection in the NIR region. *PRACTICAL SPECTROSCOPY SERIES 27*, 19-52.

30. Pahkala, K., Pahkala, E., and Syrjälä, H. (2008). Northern limits to fiber hemp production in Europe. *Journal of Industrial Hemp 13*, 104-116.

31. Pickering, K., Beckermann, G., Alam, S., and Foreman, N. (2007). Optimising industrial hemp fibre for composites. Composites Part A: *Applied Science and Manufacturing 38*, 461- 468.

32. Pickering, K., Priest, M., Watts, T., Beckermann, G., and Alam, S. (2005). Feasibility study for NZ hemp fibre composites. *Journal of advanced materials 37*, 15-20.

33. Reich, G. (2005). Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. *Advanced drug delivery reviews 57*, 1109-1143.

34. Shenk, J., and Westerhaus, M. (1991). Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop science 31*, 469-474.

35. Stevens, C., and Müssig, J. (2010). Industrial applications of natural fibres: structure, properties and technical applications, Vol 10 (John Wiley & Sons).

36. Stefano Amaducci. (2008). HEMP-SYS Design, Development and Up-Scaling of a Sustainable Production System for HEMP Textiles—An Integrated Quality SYStems Approach. *Journal of Industrial Hemp*. Volume 8, 2003 - Issue 2. Page 79-83.

37. Thygesen, L.G., Eder, M., and Burgert, I. (2007). Dislocations in single hemp fibres investigations into the relationship of structural distortions and tensile properties at the cell wall level. *Journal of materials science 42*, 558-564.

38. Toonen, M.A., Maliepaard, C., Reijmers, T.H., van der Voet, H., Mastebroek, H.D., van den Broeck, H.C., Ebskamp, M.J., Kessler, W., and Kessler, R.W. (2004). Predicting the chemical composition of fibre and core fraction of hemp (Cannabis sativa L.). *Euphytica 140*, 39-45.

39. Tourangeau, Wesley (2015), "Re-defining Environmental Harms: Green Criminology and the State of Canada's Hemp Industry", *Canadian Journal of Criminology & Criminal Justice*, 57 (4): 528–554, doi:10.3138/cjccj. 2014. E11

40. T. Randall Fortenbery, Renk Chair Professor, Michael Bennett. (2004). Opportunities for Commercial Hemp Production. *Appl. Econ. Perspect. Pol.* (2004) 26 (1): 97-117. doi: 10.1111/j.1467-9353.2003.00164.x

41. Van de Weyenberg, I., Truong, T.C., Vangrimde, B., and Verpoest, I. (2006). Improving the properties of UD flax fibre reinforced composites by applying an alkaline fibre treatment. Composites Part A: *Applied Science and Manufacturing 37*, 1368-1376.

42. van den Broeck, H.C., Maliepaard, C., Ebskamp, M.J., Toonen, M.A., and Koops, A.J. (2008). Differential expression of genes involved in C 1 metabolism and lignin biosynthesis in wooden core and bast tissues of fibre hemp (Cannabis sativa L.). *Plant science 174*, 205-220.

43. Van der Werf, H., Haasken, H., and Wijlhuizen, M. (1994a). The effect of daylength on yield and quality of fibre hemp (Cannabis sativa L.). *European Journal of Agronomy 3*, 117- 123.

44. Van der Werf, H.M. (2004). Life cycle analysis of field production of fibre hemp, the effect of production practices on environmental impacts*. Euphytica 140*, 13-23.

45. Van der Werf, H., van der Veen, J.H., Bouma, A., and Ten Cate, M. (1994). Quality of hemp (Cannabis sativa L.) stems as a raw material for paper. *Industrial Crops and Products 2*, 219-227.

46. Waśko, J., and Mańkowski, T. (2004). Applicability of Flax and Hemp as Raw Materials for Production of Cotton-like Fibres and Blended Yarns in Poland. *Fibres & Textiles in Eastern Europe 12*, 47.

47. Xu, F., Yu, J., Tesso, T., Dowell, F., and Wang, D. (2013). Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: a mini-review*. Applied Energy 104*, 801-809.

48. Zofija J, Bronislava B, Elvyra G, Jurgita C, and Ana Luisa F. (2015). Chemical composition and physical properties of dew- andwater-retted hemp fiber. *Industrial Crops and Products* 75 (2015) 206–211

Website:

https://web.archive.org/web/20110219132638/http://www.green.net.au/gf/hemp_cultivati on.htm

Appendix

Appendix 1. Biochemical and correction data of CWR%_dm

Appendix 2. Biochemical and correction data of ADL%_dm

Appendix 3. Biochemical and correction data of KL%_dm

Appendix 4. Prediction results

