TALLINN UNIVERSITY OF TECHNOLOGY

SCHOOL OF ENGINEERING

Department of Electrical Power Engineering and Mechatronics

EE70LT

ESTIMATION OF MOISTURE CONTENT OF BIOFUEL PELLETS USING HYPERSPECTRAL DATA

BIOKÜTUSE PELLETITE NIISKUSE SISALDUSE HINDAMINE KASUTADES HÜPERSPEKTRAALANDMEID

MASTER'S THESIS

MECHATRONICS PROGRAM

Student   Olabode Olaide Felix

Student Code 156358MAHM

Supervisor Märt Juurma

Tallinn, 2017

AUTHOR'S DECLARATION

I hereby declare, that I have written this thesis independently. No academic degree has been applied for based on this material. All works, major viewpoints and data of the other authors used in this thesis have been referenced.

Thesis is completed under the supervision of Märt Juurma

"......." 2017

Author: ....

/signature /

Thesis is in accordance with terms and requirements

"......." 2017

Supervisor: (Supervisor's digital signature on the next page)

Accepted for defence

"......."....................2017

Chairman of theses defence commission: ....

/name and signature/

VALIDITY CONFIRMATION SHEET

SIGNED FILES

FILE NAME	FILE SIZE
Estimation of biofuel moisture content using hyperspectral data.pdf	2.6 MB

SIGNERS

NO.	NAME	PERSONAL CODE	TIME
1	MÄRT JUURMA	38307085713	23.05.2017 16:46:14 +03:00

VALIDITY OF SIGNATURE

SIGNATURE IS VALID

ROLE / RESOLUTION

Supervisor /

PLACE OF CONFIRMATION (CITY, STATE, ZIP, COUNTRY)

SERIAL NUMBER OF SIGNER CERTIFICATE

20999548234012656015128284228664854414

ISSUER OF CERTIFICATE	AUTHORITY KEY IDENTIFIER
ESTEID-SK 2015	B3 AB 88 BC 99 D5 62 A4 85 2A08 CD B4 1D 72 3B 83 72 47 51

HASH VALUE OF SIGNATURE

30 31 30 0D 06 09 60 86 48 01 65 03 04 02 01 05 00 04 20 70 FA 53 1C 66 5F 03 5E FC 04 B0 5F E6 3D AD 80 D9 3A 50 27 39 DA D4 1D 1D 75 33 D1 69 A8 97 B1

The print out of files listed in the section **"Signed Files"** are inseparable part of this Validity Confirmation Sheet.

NOTES

TTU

School of Engineering

THESIS TASK

Student: Olabode Olaide Felix, 156358MAHM

Study programme, main specialty: MAHM02, Mechatronics

Supervisor: Early-stage researcher, Märt Juurma

Consultants: none

Thesis topic:

(in English)   Estimation of moisture content of biofuel pellets using hyperspectral data

(in Estonian)   Biokütuse pelletite niiskuse sisalduse hindamine kasutades hüperspektraalandmeid

Assignments to be completed and the schedule for their completion:

No.	Description of the assignment	Completion date
1.	Research of previous work and creation of model for the thesis	12.03.2017
2.	Planning of experiments	07.04.2017
3.	Final spectral and reference data acquisition	02.05.2017
4.	Data Analysis	12.05.2017
5.	Prediction model final test	15.05.2017

Engineering and economic problems to be solved: Designing and carrying out experiments to obtain both spectral and reference data. Analyzing the data and calibrating a model using the spectral and reference data. Using the calibrated model to predict moisture contents of biofuel pellets with their spectral data.

Language of thesis: English

Deadline for submission of the application for defense: 15.05.2017

Deadline for submitting the thesis: 25.05.2017

Student: Olabode Olaide Felix

Supervisor: Märt Juurma

TABLE OF CONTENTS

FOREWORD

This thesis topic was proposed by the Chair of Mechatronics. Research on moisture content estimation has received substantial attention in extant literature. This thesis work is an aspect of a robust solution developed by the Department of Electrical Power Engineering and Mechatronics to create a system able to measure all required parameters of biofuel pellets online without damage to the pellets or the need to take them off the production line. The measurements used for this thesis were taken at the Machine Vision Laboratory of the Department of Electrical Power Engineering and Mechatronics while reference value measurements were taken at the Department of Materials and Processes of Sustainable Energetics.

The author would like to thank the supervisor Märt Juurma for the guidance and help given all through the process of completing this work.

EESÕNA

Käesoleva lõputöö teema oli välja pakutud elektroenergeerika ja mehhatroonika instituudis mehhatroonika uurimisrühma poolt. Niiskuse sisalduse hindamisega seotud töid on ka varasemalt tehtud ning publitseeritud. Praegune töö siin aga on osa ühest robustsest lahendusest. Täpsemalt on tegu süsteemiga, mille välja töötamisega tegeleb eletroenergeetika ja mehhatroonka instituut. See süsteem oleks võimeline mõõtma kõiki vajalikke biokütuse parameetreid ilma graanuleid kahjustamata või neid tootmisprotsessidest kõrvaldamata. Mõõtmised, mis selle lõputöö kirjutamisel vajalikuks osutusid, on tehtud elektrotehnika ja mehhatroonika instituudi masinnägemise laboris. Kontrollmõõtmised on tehtud kütuste ja õhuemissioonide teadus- ja katselaboratooriumis.

Autor soovib tänada Märt Juurmad suunamise ja juhendamise eest terve töö koostamise vältel.

LIST OF ABBREVIATIONS AND ACRONYMS

AC – Ash Content

CC – Carbon Content

CEN – European Committee for Standardization

GCV – Gross Calorific Value

GLS Weighting – Generalized Least Squares Weighting

HS - Hyperspectral

HSI – Hyperspectral Imaging

LV – Latent Variable

MC – Moisture Content

MSC – Multiplicative Scatter Correction

NIPALS - Nonlinear Iterative Partial Least Squares

NIR – Near Infrared

PC – Principal Component

PCA – Principal Component Analysis

PLSR – Partial Least Squares Regression

RMSEC – Root mean square error of calibration

RMSECV – Root mean square error of cross validation

RMSEP – Root mean square error of prediction

ROI – Region of Interest

SIMPLS – Statistically Inspired Modification of PLS

SNV – Standard Normal Variate

Vis-NIR – Visible-Near Infrared

1. INTRODUCTION

1.1. General Overview

Biofuel pellets are made by grinding biomass such as wood, hay, crops, plants or plant residue into sawdust. This sawdust is then compressed through hammer mills. Biofuel pellets have quality indices that must be shown on the packaged product by its manufacturer. Apart from moisture content, which is the focus of this thesis, calorific value, ash content, bulk density and biofuel dimension are important quality indices that must be known and explicitly stated by manufacturers.

Interestingly, Estonia ranks first in the world in the production of biofuel pellets per capita. Hence, the overall objective of this thesis is to consolidate on this achievement and proffer better and faster ways of estimating the quality indices of biofuel pellets.

Hyperspectral imaging has been used for moisture content estimation in various fields, especially in agriculture. This thesis therefore aims to apply the use of hyperspectral imaging to provide a better and faster way of estimating the moisture content of biofuel pellets.

1.2. Problem statement

One of the well-established methods of estimating moisture content is the oven dry method. This is buttressed by CEN 14774, which states that the standard procedure for estimating moisture content is the oven dry method [1]. However, this method has some setbacks. The oven dry method involves taking 300 g of biofuel samples to an approved laboratory and drying at a stable temperature of 105 °C until the mass remains unchanged. Apart from this process being destructive to the pellets, it is also time wasting, as the process cannot be incorporated into an online production system, hence the need for a faster system of estimation.

As such, this study aims to use Hyperspectral imaging (HSI) to obtain both spatial and spectral information about biofuel pellets. This method gives the ability to acquire information not obtainable in spectroscopy only or in photography. HSI has been used for many different applications, ranging from cell biology to face recognition algorithms. Therefore, hyperspectral data were obtained at the machine vision laboratory of the Department of Electrical Power Engineering and Mechatronics. This was done using the Spectronon software and the acquisition camera used was Resonon PIKA II.

The approach taken for this study is to create a model using partial least squares regression (PLSR) for MC estimation. PLSR takes the spectral data, finds factors responsible for patterns in the moisture content, and then forms a model based on this prediction. PLS_Toolbox by EigenVector is a toolbox that works with Matlab. This toolbox was used for the calibration of the model. Before the model was calibrated, principal component analysis (PCA) was done on the data. Although PCA is mostly used to reduce data dimensionality, in this thesis, it was only used to observe clustering of data and to remove outliers in the data.

The findings of this study will proffer a faster system of estimating MC of biofuel pellets using hyperspectral data which will enhance production outcomes.

1.3. Objectives of the study

The purpose of this thesis is to create a model capable of estimating biofuel moisture content from hyperspectral data. This is based on the understanding that water molecules have a distinct spectral signature and the stretching of O-H bond of water corresponds to the absorption band that is common to bio-organic materials [2]. The focal aim of this study is to calibrate the model using spectral data from the visible and NIR region.

This thesis is part of a bigger project in the Department of Electrical Power Engineering and Mechatronics to create a machine vision system able to estimate the quality indices of biofuel pellets in one exposure during the production routine.

1.4. Description of tasks involved

The tasks that need to be undertaken to achieve the objective of estimating biofuel moisture content from hyperspectral data include:

I.   Review of literature to establish the best method of moisture content estimation.

II.   Creation of a model in the form of a flow chart to show the tasks to be undertaken.

III.   Setting up hyperspectral imaging environment to enhance data collection.

IV.   Acquisition of both hyperspectral and reference data to estimate moisture content.

V.   Preprocessing of hyperspectral data.

VI.   Conduction of principal component analysis to study data and remove possible outliers.

VII.   Analysis with partial least squares regression to build a model that can predict moisture content with good accuracy.

VIII.   Testing of the PLSR model.

1.5. Thesis structure

This thesis report is organized into eight chapters.

Chapter 1 provides an introduction and overview of the objectives of the thesis. The problem statement and tasks to be undertaken to achieve those objectives are explained.

Chapter 2 presents a review of relevant literature. Specifically, the development of biofuel pellets over the years and hyperspectral imaging procedures are examined and discussed. Furthermore, this chapter highlights the existing methods according to EU standards and approaches for using HSI to estimate MC. Finally, based on the literature review, a flowchart is developed to show the order of progression of tasks to be undertaken for this thesis.

Chapter 3 describes the process of acquiring spectral data, the test set up and the software used for the acquisition. The MC estimation carried out to obtain reference values using the dry oven method is also described in this chapter.

Chapter 4 discusses the data analysis procedure, result and preliminary preprocessing of the data. Also, PCA and PLSR done in the PLS_toolbox environment are also described.

Chapter 5 provides a view on future research works and recommendations based on the findings of the study.

Chapters 6 and 7 provide summaries of the thesis in English and Estonian languages respectively.

Chapter 8 lists the references used in this study.

2. THEORETICAL BASIS

2.1. Background of work and literature review

This section provides an overview of relevant literature on estimation of the moisture content of biofuel pellets using hyperspectral data. First, the history and use of solid biofuel will be discussed and the background to the study will be examined. Second, the moisture content of biofuel pellets will also be explained, and the required parameters for biofuel pellets will be discussed. Third, the existing procedures for estimating these parameters will be examined. Finally, Hyperspectral imaging, its evolution, wide applications, methods and prior efforts to use it for moisture content estimation will also be discussed.

2.2. Development of biofuel pellets

In order to understand the development of biofuel pellets and to see why they are needed, a brief overview of how they evolved over time is given. This is not exhaustive because the direction of the thesis is different, but it gives a little overview of the history of solid biofuels.

2.2.1. Firewood

Burning wood as fuel for fire has been the practice of man since the beginning of this age [3]. Firewood has been in use as the primary source of fuel before the advent of fossil fuel especially for domestic purposes [3]. Although it is not clear when humans began to use controlled fire for light and heat, it is assumed that the first fire igniting of wood must have been from lightening [4]. At a temperature of 220°C or higher temperature, most dry plant materials ignite in air to cause fire. Thus, their bioenergy is turned into heat and light [3].

Firewood is usually tied and sold in bundles. Although the use of firewood as a source of fuel has significantly reduced over the years, its use is still prevalent in rural areas of Asia and sub-Sahara Africa [3].

2.2.2. Wood chips

Burning firewood without refining it to any other form of biofuel gives the most efficient use of the bioenergy [3]. One disadvantage with firewood however, is that it is bulky. This has led to the use of wood chips which are easier to use, have smaller combustion units and are relatively more comfortable to handle [3].

Since the beginning of the 21st century, wood chips have been used for generating heat and electricity [5]. Dwellers of rural areas in developing countries find wood chippers unaffordable. Hence wood chips are not a popular choice of fuel in these areas [3].

2.2.3. Biofuel pellets

Compared to wood chips, biofuel pellets are more processed and contain less moisture content usually between 5-10% [3].

Biofuel pellets are made by grinding biomass such as wood, hay, crops, plants or plant residue into sawdust. This sawdust is then compressed through hammer mills. They are forced through 6-8mm holes of a die pelletizer to form the pellets [3].

The resulting pellets of good quality should have an average length of 3.15 – 40mm with maximum length not exceeding 45mm, <1% fine powders and should be mechanically durable [1]. Figure 2.1 below shows standard dimensions of 8 mm wood pellets with high quality.

Figure 2.1 Dimensions of 8 mm wood pellets according to EN 14961-1[1]

The smooth cylindrical shape and the small size of the pellets allow them to be fed automatically at fine calibration [3]. Also, because they are grain-like and have low moisture content and high density, they can be compactly stored and transported over long distances [3]. However, because of the high cost of pellets, their use is still limited [3].

Biofuel pellets are usually produced from sawdust and wood shavings [6], and because of the growing demand for biofuel as an energy source, there has been increased demand for alternative feedstock [6]. These alternative feedstock include cereal straws, energy crops and animal residue [6]. With the wide range of feedstock also comes the wide range of physical properties of the pellets. Biomass usually has variations in physical properties due to differences in soil type, harvest date, cultivation techniques and weather conditions [2]. These factors greatly influence the moisture content of the pellets.

Arshadi et al. found that moisture content of the raw material greatly influenced the moisture content of the pellets. They discovered that it was indeed possible to predict the moisture content of the biofuel pellet based on the raw material characteristics while taking into consideration other factors [7].

2.3. Background to the study

According to the European Committee for Standardization, CEN under committee TC335, biofuel pellets are required to be produced and packaged according to the stipulated standards for normative and informative purposes [1]. To ensure that this standard is adhered to, there are parameters given

in the standard that must be communicated to the end user about biofuel pellets. The three most important pellets parameters are the moisture content (MC), ash content (AC) and pellet dimension. Other important parameters are mechanical durability, bulk density, calorific value and ashing value [1].

The European Union prescribed methods of obtaining these parameters in the FprEN 14774 document. The moisture content is obtained by weighing the pellets before drying, and then drying the pellets to a constant mass. The weight loss is observed and the percentage of moisture content compared with the mass of the pellet [8].

Hyperspectral imaging (HSI) is used to obtain both spatial and spectral information about an object. This gives the ability to acquire information not obtainable in spectroscopy only or in photography. HSI has been used for many different applications, ranging from cell biology to face recognition algorithms.

There has been an attempt to estimate the parameters of biofuel pellets using HSI by Gillespie et al., the result of the experiment was good. With push broom line scan technique, they used near-infrared camera to obtain hyperspectral cube of pellets. The data were thereafter processed using partial least squares method to obtain the MC, AC, Calorific value (CV) and Carbon value [2].

Although not many researches have been conducted to obtain pellet parameters with HSI, there have been efforts to estimate MC of other objects, ranging from red meat to wood to the volume of water in the soil [9][10][11]. The distinct signature of water can be extracted from the HSI cube of the whole object to estimate the MC. All of these will be discussed in detail in this chapter.

2.4. Required parameters for biofuel pellets

CEN under committee TC335 is developing two important technical specifications for solid biofuels, the first deals with classification and specification (EN14961), and the second deals with quality assurance for solid biofuels (EN 15234) [1].

The pellet standards are classified into three groups namely:

- EN 14961 -1 for general use (includes pellets from different biomass raw materials)
- EN 14961 -2 for wood pellets for non-industrial use
- EN 14961 -6 for non-woody pellets for non-industrial use [1].

Though it is foreseen that the general requirements of EN 14961-1 will probably be used mainly in industry, it is meant for all solid biofuels and is targeted to all user groups [1]. Hence, following the requirements of EN14961-1 is adequate for research pursuits.

Every solid biofuel is required to have a fuel product declaration. This declaration in addition to other contents must have the exact specification of properties according to the appropriate part of EN 14961. There are two classes of specification. The first classification is the normative properties; which are mandatory to be given in the fuel specification. The most important normative properties are the moisture content, particle dimension and ash content. The second classification is the informative properties, which are optional and may be voluntarily added to the specification. An example of the informative characteristics is the bulk density [1].

The moisture content of biofuel pellets is a mandatory characteristic to be specified and it affects other characteristics of the biofuel pellets. Bridging property for instance which affects unloading and handling of the pellets is directly influenced by moisture content [12]. The moisture content also affects the net calorific value, as high moisture content will require higher heat to vaporize the moisture. The bulk density as well as the mechanical durability also known as the strength of the pellet are influenced by the moisture content of the pellet [13]. This relationship is illustrated in Figure 2.2 below:
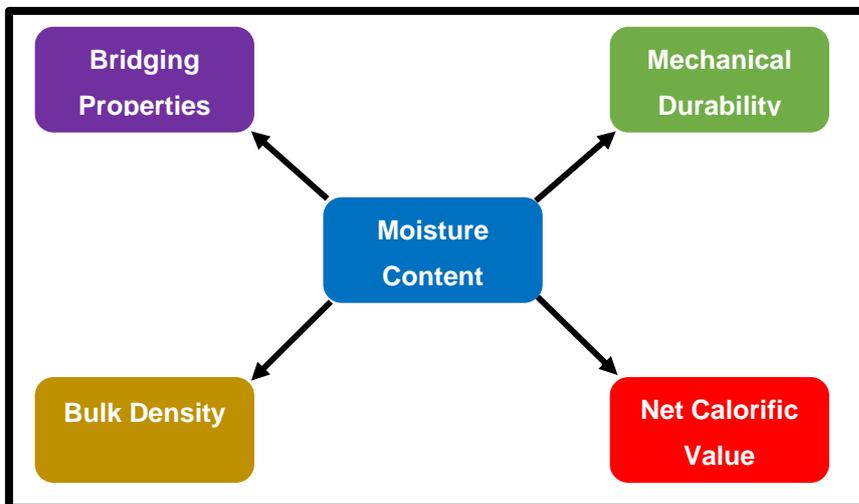
Figure 2.2 Relationship between moisture content and other properties

2.5. Existing procedure for estimating moisture content of biofuel pellets

The procedure for the determination of moisture content (oven dry method) is described in FprEN 14774. In determining the total moisture, three methods can be used.

These include: the reference method, the simplified method and moisture in general analysis sample method. The reference method is used when a higher level of precision is necessary [8].

2.5.1. Reference method (EN 14774-1)

For the reference method, a sample of minimum mass of 300g is dried in the oven at a temperature of ($105 \pm 2°C$). The air atmosphere is required to change between 3 to 5 times per hour until constant mass is achieved. From the loss in sample mass, the moisture content is calculated. Because the sample has to be weighed when still hot, it gives a buoyancy effect. This effect must be compensated for when high precision is required [14]. Figure 2.3 below shows biofuel samples in an oven.

Figure 2.3 Drying of biofuel samples [14]

2.5.2. Simplified method (EN 14774-2)

This method is similar to the reference method, except that there is no buoyancy compensation. It may be used when the highest precision is not necessary [14].

2.5.3. Moisture in general analysis sample (EN 14774-3)

For this method, the analysis sample is dried either in nitrogen atmosphere or in air atmosphere at a temperature of (105 ± 2°C). The moisture content is then calculated from the loss in sample mass. This method is applicable to all solid biofuels, and it is done in the dry oven as well [14].

2.5.4. Calculation of moisture content

Using any of the above methods, the moisture content can be calculated using Equation 2.1 below. It is expressed as a percentage by mass and should be calculated to two decimal places and rounded to the nearest 0.1% for reporting.

$$MC = \frac{(m_2 - m_3) + m_4}{(m_2 - m_1) + m_4} * 100$$

(2.1)

Where

m_1 is the mass in grams of the empty drying dish;

m_2 is the mass in grams of the drying dish and sample before drying;

m_3 is the mass in grams of the drying dish and sample after drying;

m_4 is the mass in grams of the moisture associated with the packaging [8].

2.6. Hyperspectral image acquisition applications and methods

Different molecular bonds hold biological materials together [15]. When light hits these materials, electromagnetic waves are transmitted through them.

The vibrations and bending of the molecular bonds enable spectroscopy to provide the distinct fingerprint of the materials [15].

The food industry has made remarkable use of hyperspectral imaging for quality and safety assurance of agricultural products. Visual inspection has been used over the years to assess quality and safety of products [15]. However, visual inspection is not sufficient to determine chemical properties of products. As technology advanced, machine vision techniques, such as red-green-blue (RGB) colour vision system have been used to determine external properties of products [15]. To determine chemical characteristics however, normal machine vision system is not sufficient. Hence, Spectroscopy has been used for this purpose. However, the limitation of spectroscopy is that it is inefficient in accurately analyzing heterogeneous products [15].

Repeated detection is a proposed way to solve the problem, but the setback with this solution is that more errors will be introduced. The limitations of the machine vision and spectroscopic techniques led to the development of hyperspectral imaging [15]. Hyperspectral imaging is able to obtain spatial and spectral data over ultraviolet, visible and near-infrared spectral regions (300 nm - 2600 nm) [16].

Hyperspectral imaging has a large variety of applications, which have been described in extant literature with the methods of acquisition. The applications range from clinical diagnosis of diseases [17], to face recognition which involves dealing with low signal to noise ratio, high dimension of hyperspectral data and inter-band misalignment [18]. Point Spread Function (PSF) has also been used for hyperspectral data acquisition system for ground-based astrophysical observations [19]. Push broom method was used for remote sensing. This was done by Rogass et al. who worked on the reduction of uncorrelated stripping noise in hyperspectral data acquisition [20].

Hyperspectral imaging system is made up of hardware and software parts. Although the configuration may vary depending on the object and purpose of the acquisition [15], some components are however common to the imaging system. These components include: light source; light irradiation of samples either directly or delivered by optical fiber; a detector to take both spatial and spectral information simultaneously; hyper spectrograph to disperse the wavelengths of the transmitted, reflected or scattered light and deliver signals to photosensitive surface of the detector; an objective lens, an objective table and a computer [15]. An example of configuration of hyperspectral imaging system is shown in Figure 2.4 below.
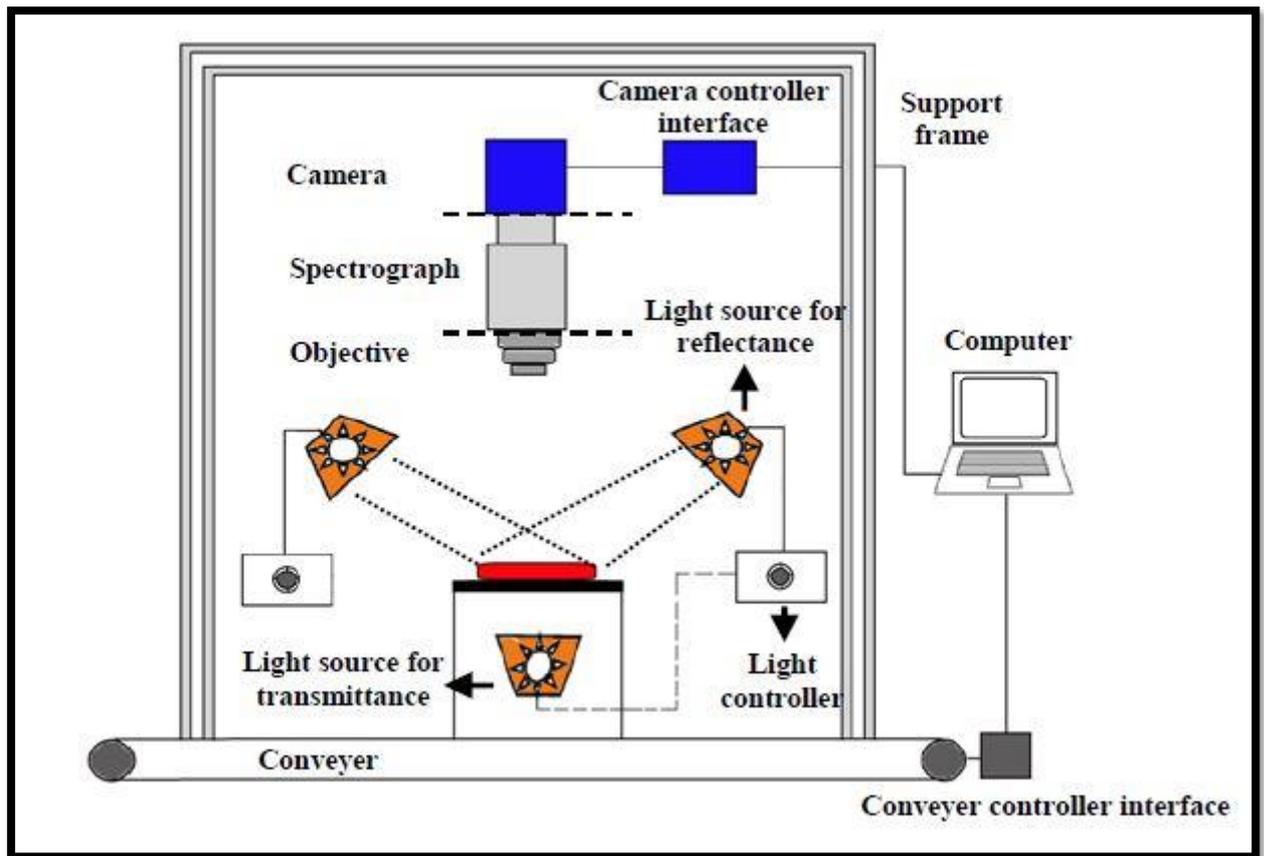
Figure 2.4 Hyperspectral imaging system configuration [15].

2.7. Application of hyperspectral imaging to moisture content estimation

Hyperspectral imaging has been used for moisture content estimation in various fields, especially in that of agriculture. Moisture content in red meat was estimated using hyperspectral imaging in the spectral range of 400-1000nm, the meat included beef, pork and lamb [9]. The moisture content of tealeaf was also estimated using HSI [21]. Kobori et al. also tested visible-near-infrared hyperspectral imaging for its suitability for monitoring the moisture content of wood samples during natural drying [10].

The most relevant literature to this research work however is that of Gillespie et al. They predicted quality indices of biomass pellet including moisture content using near infrared spectroscopy. The research findings showed good accuracy for moisture content prediction. However, the study was

not incorporated into an online monitoring system [2]. The Image was also processed with ENVI software (v4.4, Exelis Visual Information Solutions, Boulder, Colorado, USA). The work was however done using the NIR range of 880-1720 nm. At these wavelengths, they were able to observe absorbance band in the region of 1384 nm which is the first overtone stretching of O-H bonds of water molecules.

However, in the case of this thesis, a wavelength range of 420-900 nm is to be used. It is more difficult to distinctively observe absorption of water at wavelengths bands of 420-900 nm as Kamruzzaman et al. described absorption points at these wavelengths as subtle [9]. Hence, generally, the larger the range of wavelengths, the better the calibration result [2].

The goal of this thesis is to build a model using the visible and near infrared wavelengths because this thesis is part of a bigger project. This project involves incorporating the results of models to estimate other quality indices of biofuel pellets using the Vis-NIR camera available in the department of electrical power engineering and mechatronics.

A notable structure in previous studies, which was also explained in literature about predicting MC is the use of partial least squares regression (PLSR) method. The PLSR involves regressing the acquired spectral data against a reference data. The reference data in this case is the MC of each pellet sample manually estimated using an alternative method.

With this method, R^2 which shows the correlation between the predicted and the measured MC was gotten to be 0.85 while the root mean squared error of prediction (RMSEP) was gotten to be 0.727% by Gillespie et al. as shown in the figure below [2].
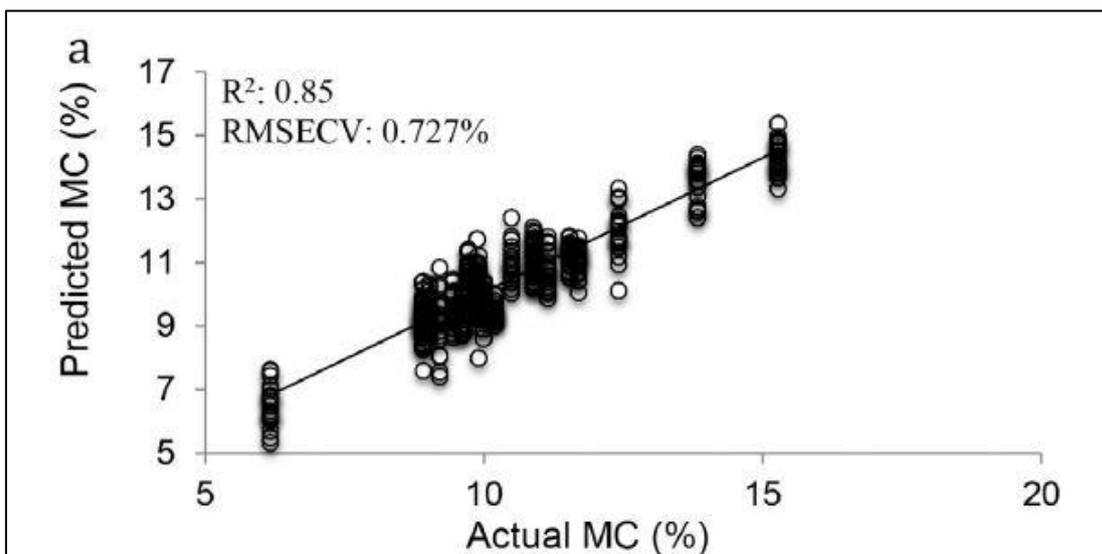

Figure 2.5 Previous result of predicted MC [2]

In order to establish the best flow of experiments, twenty-five previous studies where MC has been estimated using HSI were examined. The most important information needed are the object used for MC estimation, range of wavelength used, preprocessing technique used and the results. These are shown in Table 2.1.

Table 2.1 General overview of previous works

Reference	Wavelength range (nm)	Sample used	Model used	Preprocessing	Result
[9]	400-1000	Meat samples (beef, lamb and pork)	PLSR	ROI selection, Normalization, MSC, SNV, First and Second derivative	R^2 0,94 RMSEP 2,18%
[21]	874,41-1733,91	Longjing tea	PLSR	ROI Selection, Noise reduction (3x3 window, MNF,LOG)	R^2 0,88 RMSEP 0,211%
[2]	900-1700	Biofuel pellets	PLSR	MSC First derivative	R^2 0,85 RMSEP 0,727%
[22]	400-2500 400-750 400-1100 750-1100	Bioenergy crops	PLSR	MSC, SNV, Savitzky Golay first and second derivative	R^2 0,99 RMSCV 0,13%
[23]	780-2498	Stem and branch wood	Bi-orthogonal PLSR	N/A	$Q^2 = 0,997$ RMSEP 0,726
[24]	400-2498	Saw dust	BPLS	N/A	$Q^2 = 0,93$ RMSEP 0,53
[25]	880-1720	Beef	PLSR	Normalization, SNV, MSC	RMSEP 1,77% R^2 0,82
[26]	400-1000	Grass carp (Ctenopharyngodon idella)	PLSR	MSC, SNV	R^2 0,91
[27]	880-1720	Mango	PLSR	SNV, Mean Centering	R^2 0,995 RMSCV 2,010% RMSEP 1,408%
[28]	400 – 1000	Pork (longissimus dorsi) muscles	PLSR	N/A	R^2 0,9489 RMSEP 1,4736
[29]	430 – 960	Acid	Genetic algorithm and PLSR	ROI Selection, Mean filter, SNV	R^2 0,8162 RMSEP 5,36
[30]	900-1700	Potatoes	Partial Least Square discriminant analysis	Second derivative	R^2 0,902

Reference	Wavelength range (nm)	Sample used	Model used	Preprocessing	Result
[31]	951 -1630	Mangoes	PLSR Multivariate linear regression	Second derivative, Mean centering	R^2 0,993 RMSEP 1,282%
[32]	415 - 2501	Grasslands	OLSR, PLSR	N/A	R^2 0,25 RMSEP 0,35 R^2 0,32 RMSEP 0,19
[33]	400-1000	Pork (longissimus dorsi) muscles	PLSR	Savitzky–Golay (SG) smoothing and MSC	R^2 0,952 RMSEP 1,396 R^2 0,966 RMSEP 0,855
[34]	400 - 1000	Lychee pericarp	PLSR Successive progression Algorithm	ROI Selection, SNV	R^2 0,946 RMSEP 0,80% R^2 0,948 RMSEP 0,83%
[35]	400 – 1000 880 – 1720	Mangoes	PLSR	N/A	R^2 0,972 RMSEP 4,611%
[36]	900 – 1700	Virgin Olive oil	Genetic algorithms (GA), least absolute shrinkage and selection operator (LASSO), and successive projection algorithm (SPA	SNV Savitzky Golay	RMSCV 0,0949 R^2 0,99
[37]	400 - 1700	Iberian dry-cured ham slices	PLSR	SNV, Savitzky-Golay, Mean centering	RMSCV 1,55 – 2,31
[38]	1000 - 2500	Porcine meat	PLSR Multiple Linear Regression (MLR)	ROI selection, Masking	R^2 0,917 RMSEP 1,48%
[39]	400 – 1000	Vegetable Soybean	PLSR	ROI selection	RMSEP 4,7% R^2 0,971

Reference	Wavelength range (nm)	Sample used	Model used	Preprocessing	Result
[40]	900 - 1700	Turkey hams	PLSR	ROI selection	R^2 0,88 RMSCV 2.51
[41]	400 – 1700	Atlantic Salmon (Salmo salar)	PLSR	ROI selection	R^2 0,893 RMSEP 1,517% R^2 0,888 RMSEP 1,553% R^2 0,884 RMSEP 1,578
[42]	380 – 1100	Dehydrated prawns	PLSR, least-square support vector machine (LS-SVM) and Multiple Linear Regression (MLR)	ROI Selection	R2 0,948 RMSEP 4,394 R2 0,950 RMSEP 2,684 R2 0,955 RMSEP 2,585
[43]	350 – 2500	Soil	Linear and Multiple Regression	Savitzky-Golay	R^2 0,57 R^2 0,65

Most of the samples analyzed are agricultural produce. It can be seen from Table 2.1 that the most popular model used is the PLSR model and the result is expected to be greater than R^2 of 0,50. R^2, which is the coefficient of determination, is the parameter that shows how well the variation of the samples is explained by the model built. It is the measure of the correlation between the predicted MC and the measured MC.

It is also noteworthy that most of the studies used wavelength above 1000 nm because it is relatively easier to identify the absorption points of water at regions beyond 1000 nm [2].

The core of the predictions method is regressing the preprocessed spectral data against measured moisture content and predicting with the calibrated model.

After reviewing existing literature and establishing the best approach to take for this thesis, a flowchart model giving an overview of the tasks to be solved was developed. This covers all the processes from the selection of pellet samples to the prediction of moisture content. This flowchart is given in Figure 2.6 below:
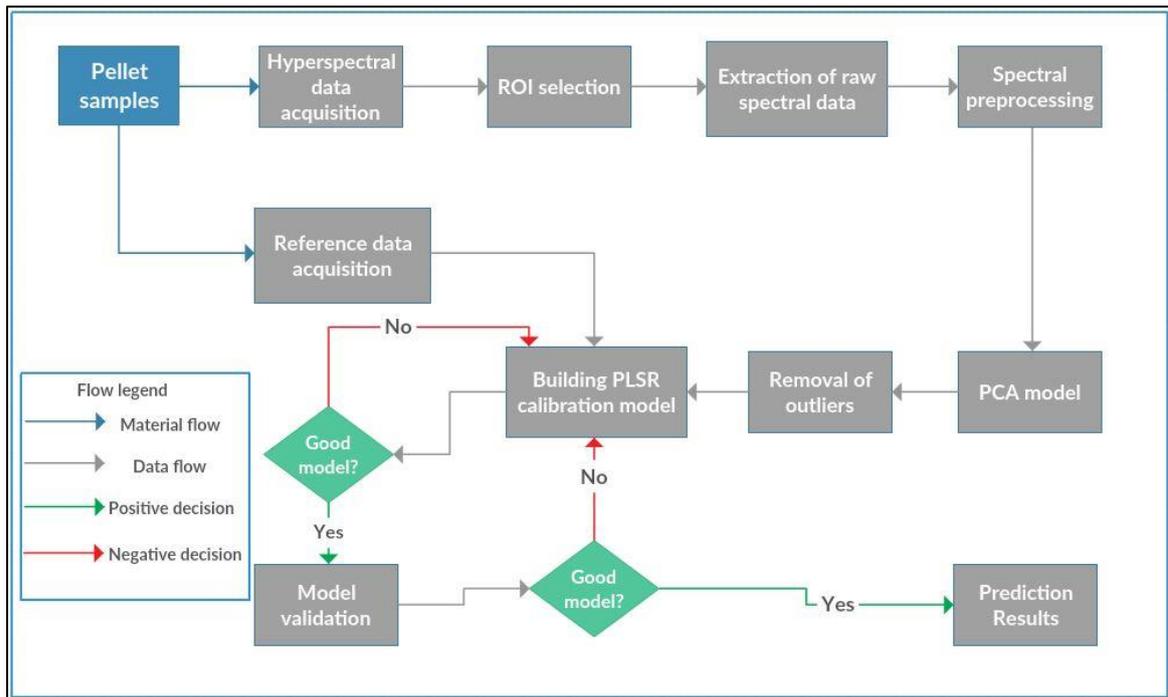
Figure 2.6 Experiment flowchart for MC prediction

The tasks listed above will be executed in the following environments.

Table 2.2 Environment for each of the main tasks

No	TASK	ENVIRONMENT
1	Sample Selection	Available samples in department
2	Hyperspectral Image Acquisition	Machine Vision Lab, Spectronon Pro Software.
3	Reference data Acquisition	Laboratory of fuel and air emission, department of energy technology, Oven compliant with CEN 14774
4	Spectral preprocessing	ROI selection on spectronon pro, other preprocessing on Matlab (PLS_Toolbox)
5	PCA model and outlier removal	Matlab (PLS_Toolbox)
6	PLSR model calibration and validation, MC prediction	Matlab (PLS_Toolbox)

3. DATA ACQUSITION

3.1. Required data for analysis

Two sets of data were needed for estimating the MC of biofuel pellets as discussed above. For acquiring the spectral data, the machine vision laboratory of the Department of Electrical Power Engineering and Mechatronics was used. The moisture content was estimated using oven dry method at the Department of Energy Technology. Both experiments are described in the sections below.

3.2. Hyperspectral data acquisition

In order to obtain the spectral data of the pellets, the Resonon PIKA II camera was used, and the datasheet of the camera is attached in appendix 1. The stage upon which the camera was mounted was fastened to aluminum profiles provided in the laboratory. A halogen lamp was used to provide lighting for the capturing. A black felt background was used to provide maximum absorption of the incident light. The choice of this material was based on the result of the study of Marshall et al. [44]. Ideally the object to be scanned should be placed on a moving stage, but this was not possible due to the limited equipment available for this experiment. Hence, the Pika II camera was mounted on a stepper motor stage.

The set up is shown in Figure 3.1 and the reasoning for the choice of this set up as well as the choice of lighting and field of view are discussed thereafter.
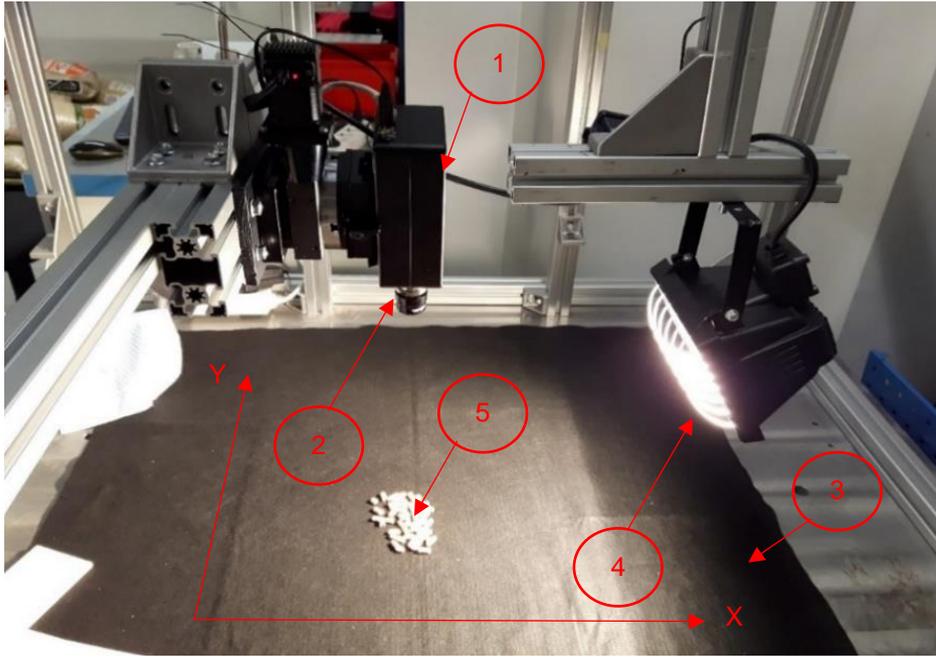
Figure 3.1 Hyperspectral data acquisition experimental set up. (1). Resonon PIKA II camera; (2). Stringray optics lens; (3). Black background; (4). Halogen lamp; (5). Test object.

The camera specifications used are described with reference to Figure 3.1. The specifications are given in the table below and they are patterned after the description used in previous projects in the department of Mechatronics [45].

Table 3.1 Camera Specification for PIKA II

Detail		Value
Lens	Name	Schneider Xenoplan 1.4/23 -0902
	Focal length	23mm
Field of view	Field of view along Y axis	103 mm
	Field of view along X axis	70 mm
Working distance		330 mm

3.2.1. Lighting and field of view

The reasoning behind the lighting source selected is to have uniform lighting. Due to the nature of the environment and camera used, halogen lamp having wavelength range of 400-2500 nm was considered. A major setback in achieving uniform lighting was that the area covered with uniform lighting is not large. The other alternative was to use two halogen lamps. This choice will however increase the heating of the stage, thereby compromising the accuracy of the MC prediction. However, since what is important to the task is the spectral data, much more than the spatial representation, the field of view was made to be as minimal as possible where enough information can be acquired while ensuring uniform lighting. Uniform lighting was achieved with field of view of 70 mm x 103 mm using a single halogen lamp. The focusing, calibration and aspect ratio setting were all done to suit this lighting choice.

3.2.2. Focusing, calibrating and aspect ratio setting

Spectronon pro software was used for both initial calibration and image acquisition. The camera was set to focus before acquiring spectral data. This was done by setting the f# level of the camera below 2.4 as specified in the user manual. The integration time was then adjusted until the brightness was below 4090 (This is also specified in the user manual). The lens was adjusted until the image of the focusing sheet obtained was well defined and clear. The lens was fixed at the best focus.

To calibrate the camera, a dark response was obtained by covering the lens with black cover to prevent light from falling on the lens. Thereafter, the reflectance was also calibrated by placing a white reference on the field of view of the camera.

The final setting was the aspect ratio setting. This was done by scanning the aspect ratio sheet iteratively and correcting the distortion by changing the frame rate and scanning speed until the correct aspect ratio was acquired. After these settings, a white reference tile was placed on the black background to observe the spectral response to the black background. This will give more insight into how the background might affect the spectral data to be obtained. The acquired image is shown in the Figure 3.2 below. It can be seen from Figure 3.2 that the black background response was just around zero brightness while the white tile is well separated.
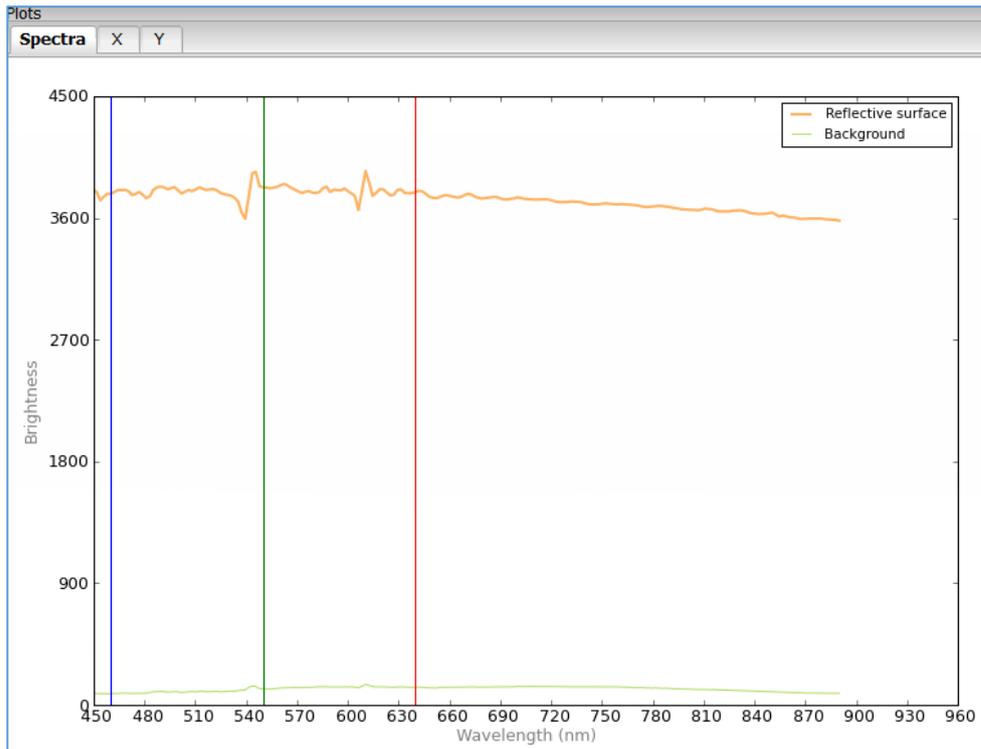
Figure 3.2 Mean spectra of white reflective tile and the black background.

While setting the aspect ratio for the acquisition, other important parameters were also set. These include: the camera setting and the stage setting (The camera is mounted on the stage). These parameters are given in Table 3.2.

Table 3.2 Acquisition parameters used

Parameters		Value
Camera settings	Frame rate (hz)	92,0
	Integration time (ms)	5,3
	Gain (dB)	19,4
Stepper motor settings	Angular velocity (deg/s)	1,35
	Homing speed (deg/s)	23,1
	Jog Speed (deg/s)	22,3

3.2.3. Data capturing

Biofuel pellets were grouped into 10 classes of 300 g each which is in accordance with EN 14774. In anticipation of the oven dry method of MC estimation, each of these classes were scanned with

the HSI camera and the acquisition file was saved in form of .bil file format. It was also necessary to select regions of interest by using the spectronon pro software and saving their mean spectrum. Ten regions of interest were identified on each set and their mean spectra were extracted. Spectral binning of 3 was initially used for the first sets of experiments in order to reduce data dimensionality. However, because the mean spectra are the data to be mostly used, the binning was removed and rather than having 80 wavelength points, 240 wavelength bands were used. Figure 3.3 shows the spatial and spectral data obtained for one class.
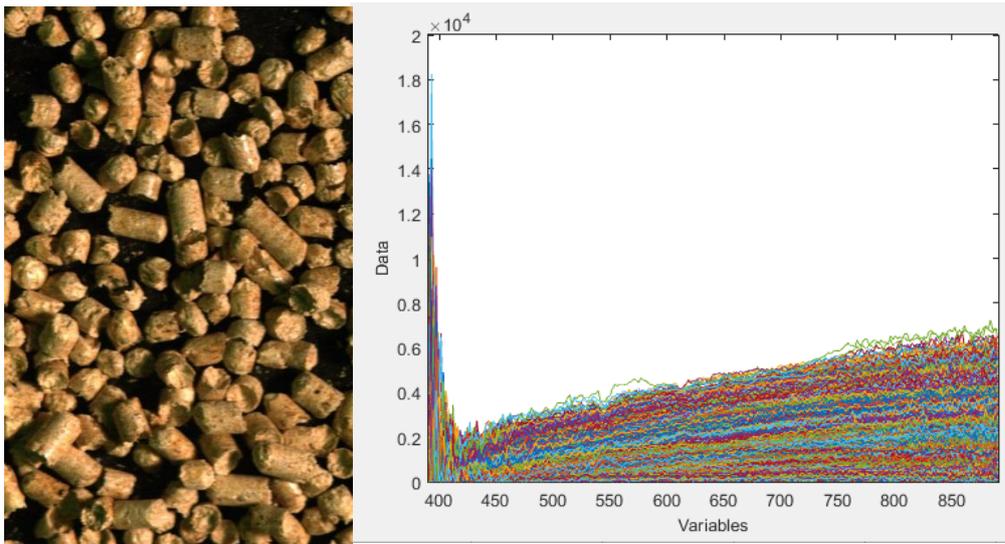
Figure 3.3 Acquired data: Spatial image (left) and hyperspectral data (right) of pellet set seven

In quick succession of approximately 15 minutes interval after obtaining a reliable hyperspectral data, the pellets were taken to the laboratory of fuel and air emission analysis at the department of energy technology for MC estimation.

3.3.   Reference data acquisition

The reference data are needed for the calibration of a PLSR model. To acquire this data, there is a standard that must be followed. The FprEN 14774-2 gives the specifications of equipment to be used and the procedure that must be followed in using the oven dry method to estimate MC.

The pellets are to be taken to the laboratory in sealed air-tight containers or bags. Test portions are required to be of nominal size 30 mm prepared in accordance with CEN/TS 14780. The drying oven

must also be capable of being controlled at a temperature within the range of (105 ± 2) °C in air atmosphere until constant mass is achieved. The air atmosphere is required to change between three and five times in an hour. Figure 3.4 shows the packed samples for MC estimation.

Figure 3.4 Samples taken for MC estimation

The Gallenkamp hotbox oven size 2 was used for the experiment, the specification of the oven is given in Table 3.3.

Table 3.3 Specification of Gallenkamp hot box size 2 oven [46].

Parameter	Value
Temperature range (°C)	40 to 200
Maximum power rating (W)	600
Fluctuation (°C)	± 1,5

The trays were weighed when empty and also when the samples have been placed in them. Pellets were allowed to dry for an hour. They were weighed and the weights were recorded until constant mass was achieved. The arrangement of the pellets was in such a way as to allow free circulation of air in the oven. This is shown in Figure 3.5 below.

Figure 3.5 Arrangement of samples in Gallenkamp oven

After three hours of drying, the weight of all the samples became constant and were all recorded. Equation 2.1 was used to calculate the MC from the weights. Table 3.4 below shows the weights and calculated MC. MC of biofuel pellets are given in percentage of their mass written as (w-%), they are to be calculated to two decimal places and rounded to the nearest 0,1 % [8].

It is worth mentioning that not all the samples attained constant weight at the same time. Some of the samples reached constant weights after two hours while some others took three hours. The lowest weights were used for samples that began to re-moisturize.

Table 3.4: Result of MC estimation

Sample	Mass of drying dish (m1) g	Dish + sample before drying (m2) g	Dish + sample after drying (m3) g	$MC = \dfrac{m_2 - m_3}{m_2 - m_1} * 100$ (MC) $w - \%$
Wood pellet 1	467,5	767,4	745,8	7,2
Wood pellet 2	465,8	766	744,4	7,2
Wood pellet 3	466,6	766,4	744,6	7,3
Wood pellet 4	472,1	772,2	751,1	7,0
Wood pellet 5	467,9	768	746,8	7,1
Wood pellet 6	470,8	770,5	750,3	6,7
Wood pellet 7	465,9	765,8	744,4	7,1
Wood pellet 8	471,4	771,3	749,6	7,2
Wood pellet 9	229,2	529,5	509,3	6,7
Wood pellet 10	227,9	527,8	507,9	6,6

These values of moisture content were used as reference values for calibration and validation of the PLSR model.

4. ANALYSIS AND TEST RESULTS

4.1. Overview of the analysis done

This chapter describes the analysis of the data obtained in chapter 3 and the description of the test result for the prediction of MC. This is in accordance with the description given in chapter 2. The chapter is divided into four sections. These include: data preprocessing, principal component analysis, partial least squares regression and prediction results. The analysis and prediction result were majorly done with Matlab R2015a using Eigenvector PLS Toolbox. The initial preprocessing was done in Spectronon pro environment – the same software used for the HSI data acquisition.

4.2. Preprocessing

One of the aims of the preprocessing is to isolate the needed data from the chunk of data obtained. HS data are quite huge, and attempting to run analysis on the entire data does not only take a lot of machine's memory space but it also introduces many errors into the analysis. Different variations in data are usually captured which include: variation due to lighting, camera error at the start of image acquisition and before homing. To take care of these errors, several preprocessing tools are available and some of these have been successfully used in extant literature. However, the best preprocessing techniques can only be ascertained by experiments. This is because experiments for which the preprocessing technique was suggested may vary considerably from the present experiment.

The preprocessing techniques that were used in this task are ROI selection, MSC (mean), Mean centering, Autoscale, Standard normal variate, Savitzky Golay second derivative filter, GLS weighting, Normalization and Orthogonal signal correction. The ROI selection was done on Spectronon software while the rest of the preprocessing were done on PLS_Toolbox. On PLS_Toolbox, two main analysis were done (PCA and PLSR). For PCA, MSC, savitzky golay and mean centering were applied in that order. For PLSR on the other hand, SNV, GLS weighting and mean centering were applied respectively for the calibration while normalization, orthogonal signal correction and autoscale were used for the recalibration of the model. It is expedient to explain the purpose of each of these preprocessing techniques to indicate the reasons for their selection at different phases of the task. The sections below describe the purpose of these techniques and the effect they had on the spectral data obtained.

4.2.1. ROI selection

In acquiring the HS data of the pellets, there were some broken particles of the pellet on the black felt background. The main aim of the ROI selection was to isolate solid pellets from both the crumbled ones and the background.

The Spectronon pro software was incorporated with ROI selection tool. This tool was used to select ten ROIs in each set of pellets and the aim of this is to obtain hundred ROIs in all. The mean spectra of each ROI was obtained and these were all exported as text files.

Figure 4.1 below shows the ROI selected and the mean spectra for a class of sample.
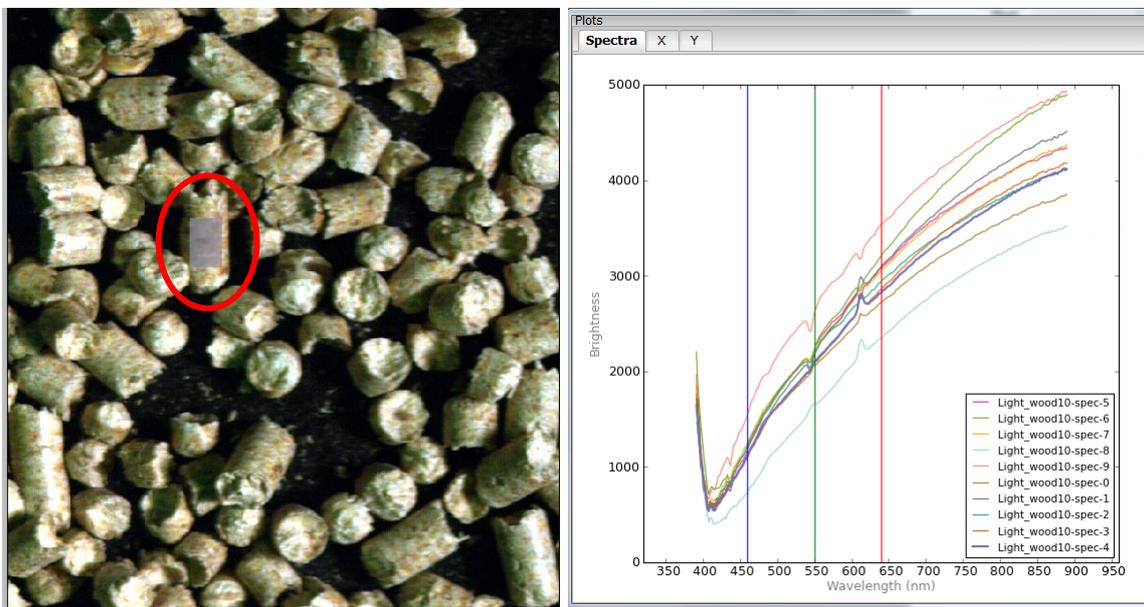
Figure 4.1 Selection of ROI: an ROI - circled in red (left), mean spectra of ten ROI (right)

All of the mean spectra were combined into a matrix by transposing the data in Microsoft Excel with each row representing the mean spectra of a specific sample and every ten rows representing a set of samples with identical estimated MC. The resulting data was loaded into PLS_toolbox for further preprocessing. The loaded data contained a measure of irregularity in the form of noise at wavelengths below 420 nm. This noise has been observed in previous studies and is generated by the camera at the start of acquisition process. This is unwanted and it may greatly distort the accuracy of the model being calibrated. Hence, this part was excluded from the spectra data by selecting the needed wavelengths in the plot window and excluding the other parts. This is shown in Figure 4.2.

36
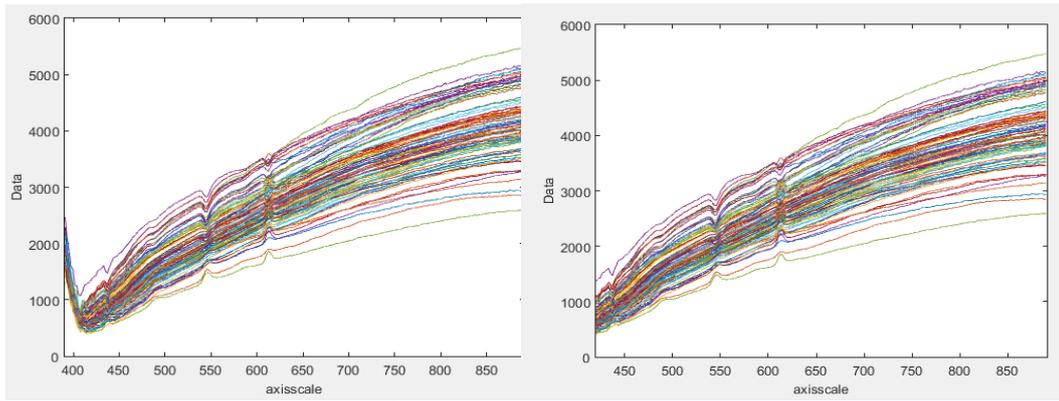
Figure 4.2 Removal of noisy wavelengths: Spectra with noisy wavelengths (left), cleaned up spectra (right)

4.2.2. MSC (Mean)

Multiplicative scatter correction (MSC) was used before the PCA analysis. It is a technique used in HSI to correct signals that contain some level of noise. It is ideally used to correct areas of spectrum that contain no chemical information, but because these areas do contain high noise to signal ratio, the MSC is usually applied to the whole spectrum. What the MSC does is that it sees the scattering in signal –the light scattering or change in path of the spectra, and fits it to an ideal spectrum that is estimated as a sample.

The result is that the spectra are made to look like they have similar patterns. This has the possibility of distorting chemical properties of materials especially when the samples are of different chemical compositions. However, because all the pellets combined into this spectrum have similar chemical compositions, this preprocessing technique is justified, as the spectra would be fitted into a sample that fits the path of biofuel pellets acquired.

Applying this preprocessing technique gave better results than the standard normal variate, probability quotient normalization and normalization techniques for the PCA analysis.

Figure 4.3 below shows the effect of applying the MSC (mean) to the acquired spectral data.
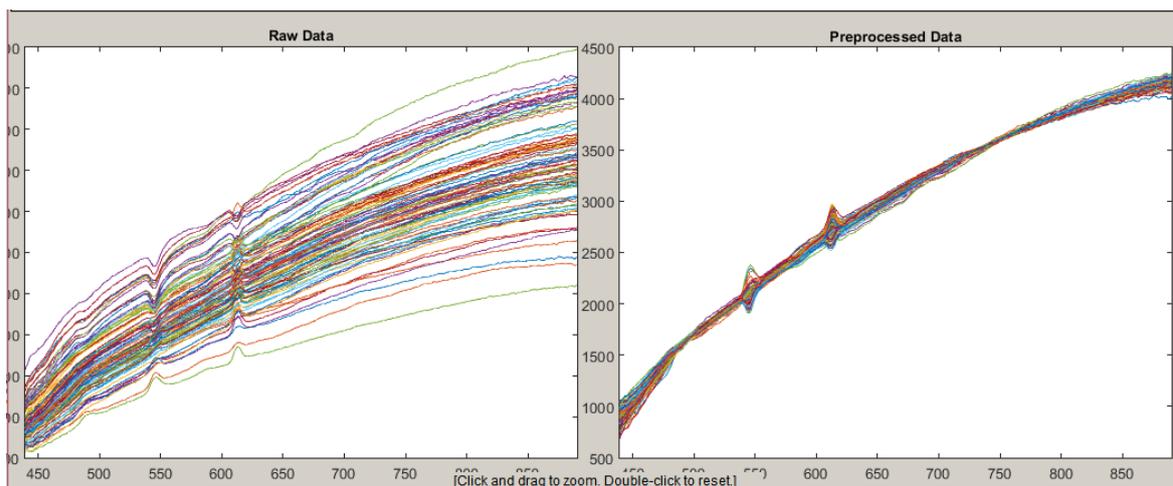
Figure 4.3 Effect of MSC: Raw data- (left), preprocessed data with MSC (Mean)- (right)

The absorption and peak points are retained as seen in Figure 4.3. This is important, as it is essential for the PLSR process of identifying water's O-H stretching.

SNV is another type of scatter correction similar to MSC, it is the second most widely used scatter correction [47]. In this work, the effect of SNV on the calibration data was similar to that of the MSC. However, the MSC gave better PCA model while the SNV gave better PLSR model. Hence MSC was used for PCA while SNV was used for PLSR model.

After the MSC, savitzky golay and mean centering were used before PCA analysis. The purpose and structure of savitzky golay are described in subsequent item.

4.2.3. Savitzky golay (second derivative)

Random noise and slow variation usually corrupt measured spectra. It was observed as earlier shown that the beginning and the ending of the acquired spectral data contained some level of noise. A major advantage of savitzy golay is its ability to identify absorption peaks in acquired spectral data. This was indeed the purpose for which the filter was designed [48]. In this work, the absorption peaks are central to identifying the presence of water and preserving this in the analysis is good for the model being calibrated. In the acquired data, some level of noise were observed. Though this was not much, the savitzky golay filter helped in filtering out the noise. The filter was used before PCA and it greatly improved the result obtained.

4.2.4. Normalization

This preprocessing technique corrects scaling and gain effects that vary across different samples. In spectroscopy, the causes of these effects include path length effect, scattering, detector variation and other instrumentation sensitivity effects [49]. The normalization preprocessing helps to give the same impact to all samples allowing them to contribute equally to the model.

4.2.5. Mean centering

This preprocessing technique was used for both the PCA and the PLSR in this task. Mean centering is one of the most common preprocessing techniques and the most common centering method. It takes every entry in a column and subtracts it from the average of the column. The effect is that the resulting data shows how different the entry is from the average of the original data.

This preprocessing is usually required for PCA and PLSR. This is because mean centering eliminates the need for an intercept from the PLSR model, since lesser terms of the regression model need to be estimated. Estimation may be more precise after mean centering the data. However not all mean centered data yield results with good precision. Hence, both mean centered and non mean centered data must be tested to know the best result [50].

The effect of mean centering on the spectral data with MSC already applied is demonstrated in Figure 4.4
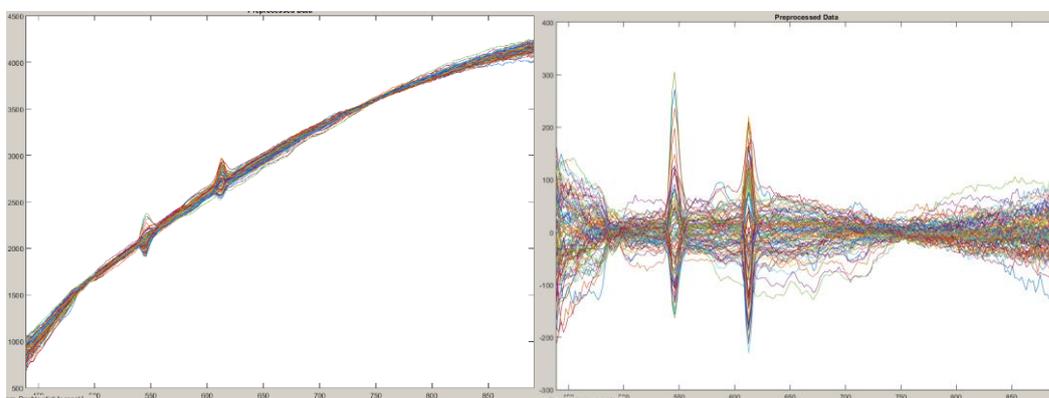

Figure 4.4 Effect of mean centering: (left) non-mean centered, (right) mean centered.

4.2.6. Autoscale

This method uses the mean-centering and then divides each column by the standard deviation of the column [49]. This method is used if the source of the variation is the signal and not the noise. Hence when this technique is used, it is better to use it at the end of the preprocessing steps, after correcting the noise with other preprocessing techniques. This method was used in place of mean centering for some of the experiments as will be discussed later.

4.2.7. Generalized least squares weighting (GLSW)

GLS weighting is a filter gotten by finding the differences between samples which should otherwise be similar [51]. These differences are seen as interferences and as such are down-weighted. Simply put, when it is applied in regression, it checks the Y block and then filters out from the X block, sources of variance not caused by the Y block. This is essential in multivariate data because several factors are responsible for variations in the spectral data. This filtering allows the regression model to achieve better calibration and prediction. There is an adjustable parameter in GLS Weighting (α), when it has larger values, the effect of the filter is less and when it has lower values, the effect of the filter is more. The GLSW was used to improve the result of the calibration and consequently of the prediction.

4.2.8. Orthogonal signal correction (OSC)

The OSC is similar to the GLSW in that it tries to optimize the variance caused by the Y block. However, in achieving this, the OSC removes the variances from the X block that are orthogonal to Y block. It identifies the principal components in X block and then rotates the loading to make them orthogonal to the Y block. These are the scores not influenced by the Y block data. After this, a PLS model is built to predict the orthogonal scores. This model is eventually used to remove the orthogonal components. This method was intended to make improvements on the GLSW predictions.

4.3.   Principal component analysis (PCA)

Principal component analysis (PCA) is a popular method used when dealing with large volume of data. Because of the large volume of the data, it is difficult to use graphical illustrations to understand the nature of the data. Therefore, PCA is able to find relationships between this data by establishing the major axes of the clustering of the data. A major use of PCA is the reduction of data dimensionality. PCA identifies patterns in data and it then expresses the data in such a way that the similarities and the differences are highlighted. When these patterns have been found, the data can then be compressed by reducing the number of its dimension without damaging the information. Although some details may be lost, important information is largely preserved.

PCA was not used in this task to reduce the dimension, but to observe clustering of the data and to remove any possible outliers that may affect the quality of the model being built. This section gives a description of PCA and explains how it was implemented on the spectral data. The result is also discussed. The MatLab R2015a toolbox PLS_Toolbox was used for this PCA analysis.

The PCA was done using Venetian blinds cross validation. The maximum number of principal components (PCs) was set at 20, the number of data splits at 10, and one sample per blind which is also referred to as the 'thickness'.

Cross validation explains how well the data fits into the created model and how other similar data not used in the calibration will fit as well. Cross validation also helps in choosing the number of PCs for the model calibration. The number of PCs chosen ultimately determines the accuracy of the model. As a rule, the number of PCs are chosen in such a way that the root mean square error of calibration (RMSEC) and root mean square error of cross validation (RMSECV) are least. A PC can also be chosen if the RMSEC and RMSECV improves in the next PC but the improvement is not significant. PCA was done for the pre-treated spectral data, and the choice of the PC is shown below.

Table 4.1 Percentage variance, RMSEC and RMSECV of the PCA model

| No of PCs	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative	RMSEC	RMSECV
1	1.04e+03	60.00	60.00	1.779	2.215
2	3.11e+02	17.92	77.92	1.322	1.922
3	1.63e+02	9.39	87.31	1.002	1.738
4	8.10e+01	4.67	91.98	0.7969	1.478
5	3.67e+01	2.11	94.09	0.684	1.547
6	2.55e+01	1.47	95.56	0.5931	1.529
7	2.17e+01	1.25	96.80	0.5031	1.319
8	1.26e+01	0.72	97.53	0.4426	1.454
9	7.58e+00	0.44	97.96	0.4017	1.526
10	6.66e+00	0.38	98.35	0.3619	1.636
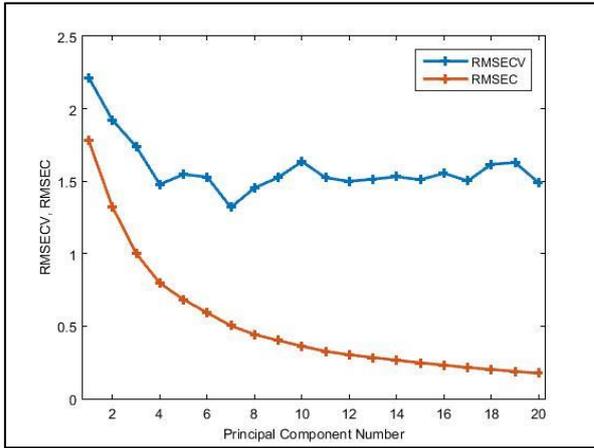
Figure 4.5 RMSECV and RMSEC vs Number of PC

Table 4.1 and Figure 4.5 show the PCs with their percentage of variance captured, the cumulative variance, RMSEC and RMSECV. The PLS_toolbox software suggested selecting seven numbers of PCs; this follows the explanation given above. It can be seen that below seven PCs, the RMSECV increased and if more PCs are selected, this will introduce more error into the model. Also at this point, 96,8% of the variance has been captured by the first seven PCs. The first PC captured 60%. It can also be seen from the RMSECV curve that this point is the lowest. The change in RMSEC from this point is also minimal.

After reviewing this plot, the other plots that give the required information for this task is the Q residuals, hotelling T^2 and the score plots. The Q residual and hotelling T^2 plots show how the samples are similar to one another and how they are different from one another. The score plot also provides information about each of the samples. Furthermore, it is actually on the score plot that one can observe this clustering the best and identify possible outliers.

These plots are given below and an explanation of the plots given thereafter.
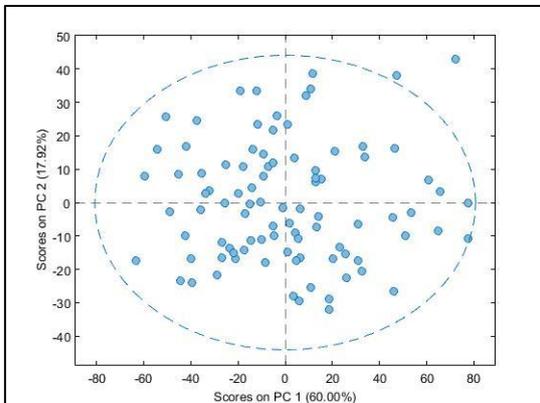
Figure 4.6 Plot of scores of PC2 against PC1
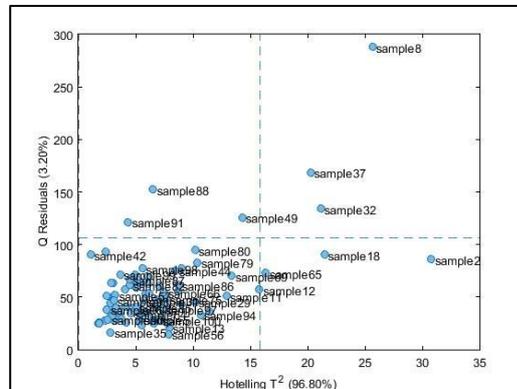
Figure 4.7 Plot of Q residuals against hotelling T^2

42

It can be seen from Figure 4.6 that all but two of the samples fall within the circle, and this circle represents the 95% confidence limit. This is the level at which the model can explain the relationship between samples. There are two samples outside the 95% confidence limit, although one is close to this limit, the other is way off the limit.

The Q residuals and Hotelling T^2 plot give some information about the similarity of the model. Again, two samples (Sample 8 and Sample 2) are very dissimilar to other samples, although sample 2 has high Hotelling T^2 value, its Q residual value is low. However, sample 8 has both high Q Residuals and Hotelling values.

In other to take a good decision about these points, it is important to check them again on the score plot while colouring them with their Q residuals value. It would not matter if in spite of their high Q residual and hotelling values, they still fall within the 95% confidence limit. Hence, the plot with this colour is given in the figure below.
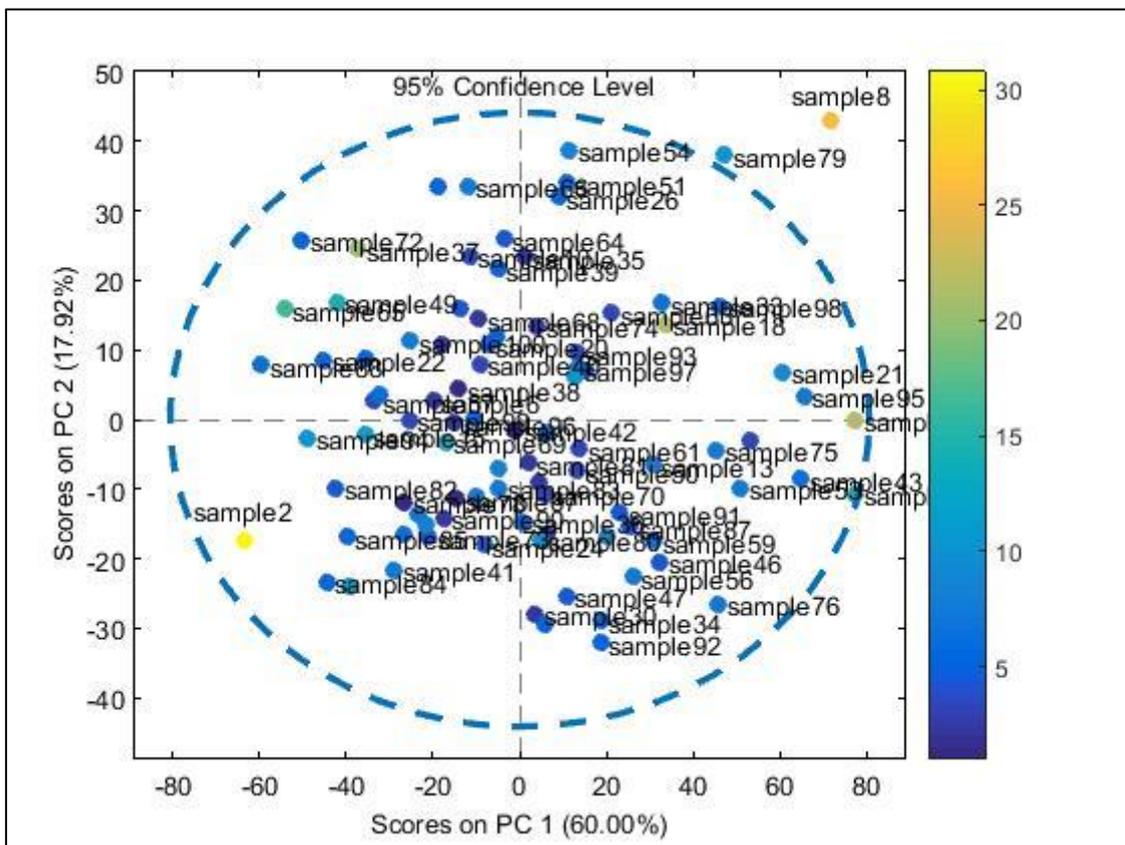
Figure 4.8 Score plot with samples coloured with their Q residual values

From Figure 4.8 it can be seen from the top right that sample 8 is indeed the sample that falls out far from the confidence limit. From the lower left corner, sample 2 is well explained by the model as it falls within the confidence limit of 95%.

Hence, since all the pellets were in the same package and ideally should be quite similar, the sample 8 with high hoteling and Q residual value is considered as an anomaly and an outlier. This sample was removed from the data set and the model was recalibrated.

The new calibrated model performed better. The first seven PCs covered 97,09% variance with RMSEC value of 0,4909 and RMSECV of 1,298. These values are better than those of the previous model which captured 96.80%, 0.5031, 1.319 of the respective parameters. The previous sample 79 immediately outside the confidence limit now falls within the confidence limit as shown in Figure 4.9.

PCA is reliable in identifying and removing outliers. It has been shown in literature that this process increases the accuracy of the model and removing a data sample through PCA does no harm to the model being built [37].
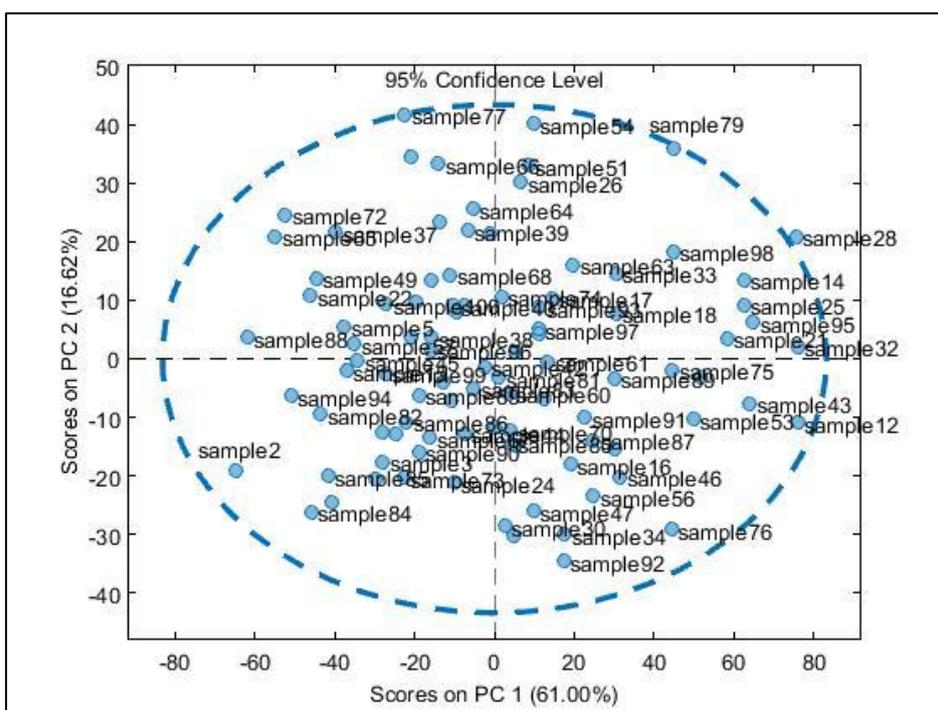
Figure 4.9 Score plot of the recalibrated model

Because PCA uses linear transformation for the decomposition of spectral data, it forms several components that are uncorrelated which makes the reconstructed data from the model not ideal for PLSR [52]. However, after removing the outliers, the original data was retained and not the principal components.

The new data from the PCA with the outlier removed was then used for PLSR which is the main analysis to create the model for MC prediction. This is described in the next section.

4.4. Partial least squares regression (PLSR)

Partial least squares regression (PLSR) is a multivariate regression method [53] that takes a given number of components to predict the independent variable 'y' from dependent variable 'x' [51]. In order to use PLSR to predict MC, the data with the outliers removed from the PCA was used for this analysis. The first set of preprocessing techniques applied are the MSC (mean) and the mean centering. The SIMPLS algorithm, which is the default of the software, was used, other algorithms such as NIPALS gave similar results for this analysis. Figure 4.10 below shows the PLS_Toolbox interface used for PLSR and the explanation of the each of the steps taken to obtain the final predicted value are also discussed. It is worth mentioning that this interface is similar to what is used for the PCA analysis. However the difference is that the independent variable 'Y' used for PLSR is not used in the PCA analysis.

	X-Block LV	X-BLock Cumulative	Y-Block LV	y-Block Cumulative	RMSECV Moisture	
1	49.88	49.88	2.41	2.41	0.24224	
2	20.70	70.58	4.74	7.15	0.24285	
3	14.25	84.83	2.98	10.13	0.2413	
4	6.09	90.92	5.31	15.44	0.23878	
5	4.49	95.41	9.42	24.86	0.23987	
6	1.23	96.64	23.67	48.53	0.22858	current*
7	0.88	97.52	6.38	54.91	0.22573	
8	0.22	97.75	12.08	66.99	0.22134	
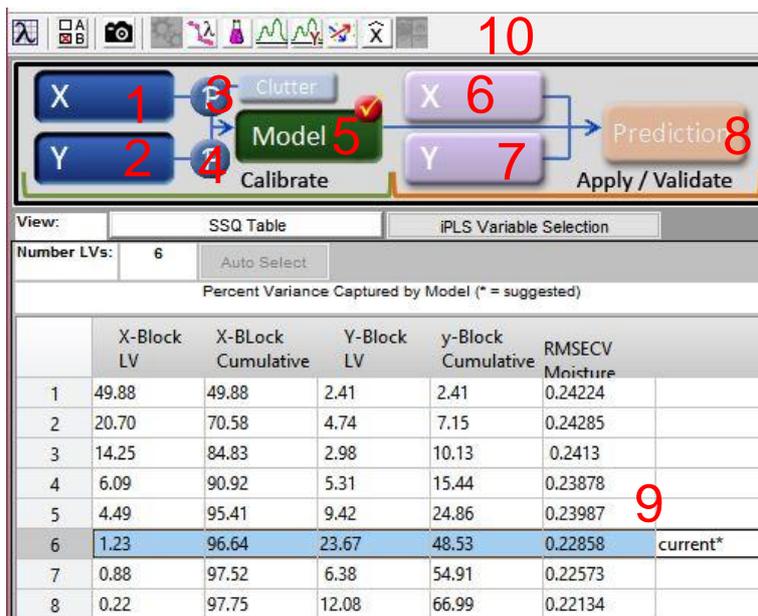
Figure 4.10 PLS_Toolbox Analysis Interface

In Figure 4.10 the main icons are labelled in red, these are used for the analysis as explained below:

1. X block data can be loaded and edited in 1. In the case of this thesis, the x block data is the spectral data with the outlier removed using PCA analysis.
2. The Y block data is loaded in 2, the Y block data is the measured moisture content from the laboratory. It is mandatory for the Y block data to have the same number of rows as the X block data; otherwise, the cross validation and prediction cannot be done.

3. This is for the preprocessing of the X block data. For the first set of analysis, MSC and mean centering were selected.
4. For PLSR, the Y block data also needs to be centered and mean centering was used for the Y block data.
5. After preprocessing, the model is calibrated using this button and a calibrated and cross validated model is created. For this task, the venetian blinds cross validation method was used, with maximum numbers of LVs at 20, number of data splits at 10 and the thickness at 1.
6. The validation/test spectral data is loaded into 6. It is with this data that the prediction is made, which is the main result of this task.
7. The measured MC corresponding to the loaded spectral data is loaded into 7. The accuracy of the prediction depends on how close the predicted data is to this data.
8. This generates the prediction based on the loaded data.
9. The number of Latent variables is selected here based on the value with the least error of cross validation.
10. Here different plots that give more information about the model can be generated.

The number of latent variables which was selected was 6 as it covered 96,64% of the data and it has the least amount of root mean squared error of cross validation, as shown in Figure 4.11 below:
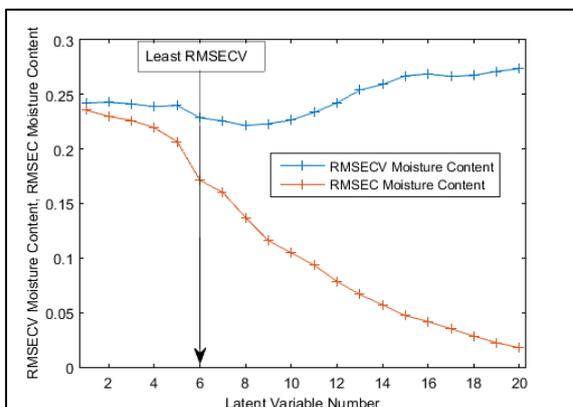
Figure 4.11 RMSEC and RMSECV against LV

The result gotten with the MSC filter was poor with coefficient of determination R^2 of calibration at 0,48%. Since there is no standard rule to follow in selecting preprocessing technique, the MSC was changed to SNV. After this modification, even though the RMSEC and RMSECV were good at 0,13675 and 0,22108 respectively, the coefficient of determination R^2 was 0,678 for calibration and 0,239 for cross validation as shown in Figure 4.12.

R^2 is a measure that shows the correlation between the predicted and the measured variable and explains how well the variance of the points can be explained by the model.
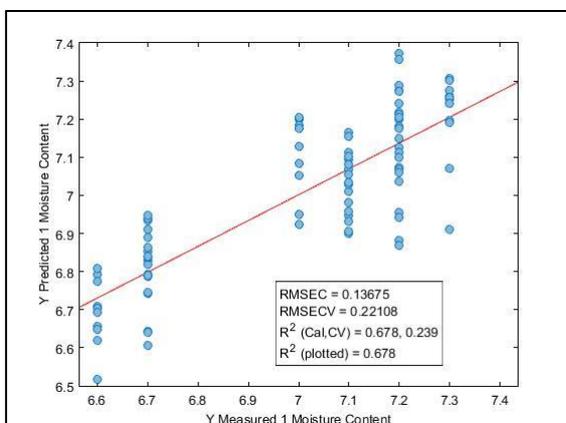
Figure 4.12 Predicted moisture content against measured moisture content

Usually in PLSR, there are possibilities that other factors such as variation in lighting and distortion due to camera movement have contributed to the prediction. Hence, a way to make sure that only the moisture content contributes to the prediction is to use the GLS weighting preprocessing technique.

This was applied and the result significantly improved. The parameters used for this GLS weighting preprocessing are given in Table 4.2 GLS weighting parameters below:

Table 4.2 GLS weighting parameters

Declutter Settings	
Clutter source	y-block gradient
Algorithm	GLS weighting
Declutter threshold	1,3

This improved the result significantly with the correlation between measured moisture content and coefficient of determination R^2 of cross validated moisture content increasing to 0,71 and that of calibration increasing to 0,99. This calibrated model is fair, considering the fact that with the limitation of few sample types and reduced wavelength range in Vis-NIR range comes the increased error. However, for the purpose and scope of this study, this result is sufficient. The results of the calibration are shown in figures below. The variance-captured table is provided in appendix 5.
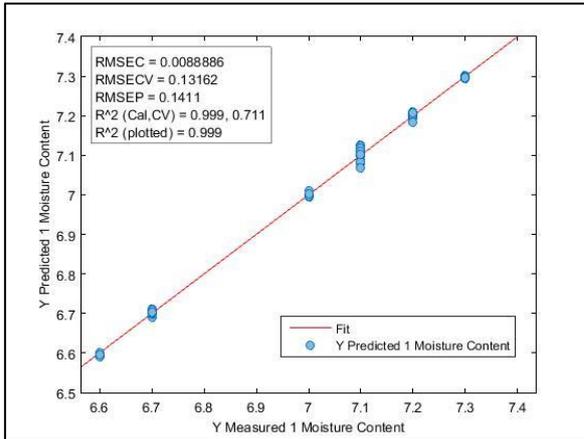
Figure 4.13 Predicted moisture content against measured moisture content
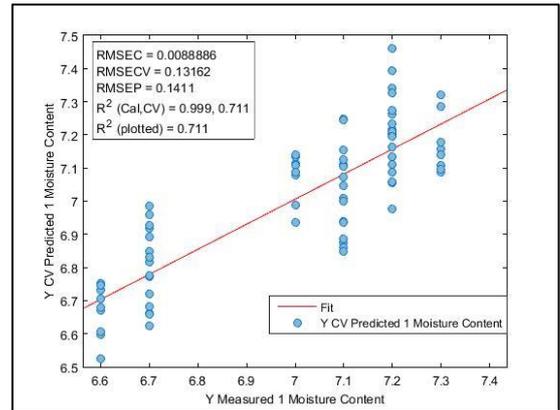
Figure 4.14 Cross-validated moisture content against measured moisture content

The figures above show the result of the calibration of the PLSR model. This model's correctness is pivotal to the accuracy of the prediction since the calibrated model in itself must have good accuracy in term of its cross validation and must also give a good result for the predicted MC.

The result of the RMSEC in Figure 4.13 is notably different from RMSECV and RMSEP. This is because of the limited number of samples used for the calibration. This value may be higher with an increase in number of samples because with the increase in sample number also comes the possibility of increase in the accuracy of the model. Increasing the quantity of samples was not considered in this thesis because all the available pellets for the work have similar moisture content. Also, there is limited space in the laboratory for fuel and air emission analysis for the purpose of estimating reference data.

This model was saved, and it was tested using 40 ROIs selected in a similar manner to what was described in item 4.2.1. The testing of the model is described in the section below:

4.5. Testing of the PLSR model

Using the calibrated model and the selected validation spectral data with their corresponding measured MCs, the model was tested by loading the testing data to the toolbox. The model predicted the MCs of the spectra with R^2 of 0,732 and RMSEP OF 0,1411.

The result of the prediction is shown in Figure 4.15 below.
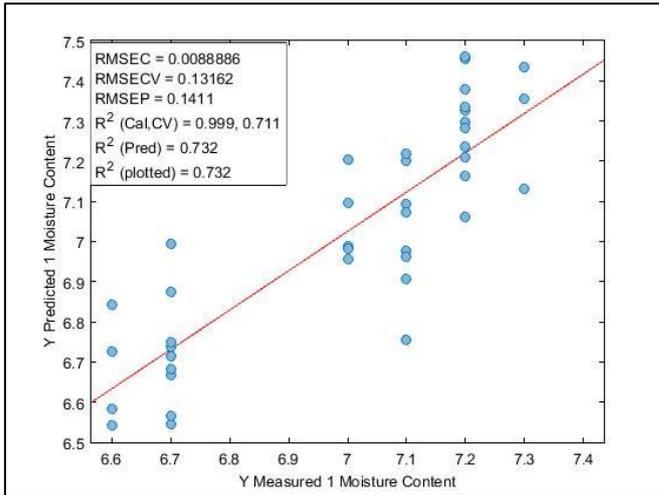
Figure 4.15 Predicted MC against measured mc

In literature, this prediction result is only considered fair and not a good model. Hence, there is a need to recalibrate the model to obtain at least a good prediction.

Table 4.3 below shows different values of R^2 and their interpretation according to Fagan et al. [22].

Table 4.3 Interpretation of coefficient of determination R^2

R^2 Value	Interpretation
0,50 – 0,65	Discrimination can be made
0,66 – 0,81	Approximate quantitative predictions
0,82 – 0,90	Good predictions
>0,91	Excellent predictions

In order to get the coefficient of determination of prediction to at least a good prediction, the model was re-calibrated. This is in accordance with the flow of experiment laid out in Figure 2.6.

4.5.1. Recalibration of PLSR model

For the new model, new preprocessing methods were selected. Again, there is no hard and fast rule to selecting preprocessing techniques, so they are selected iteratively based on the understanding of their purpose and effects as previously explained in Section 4.2.

The normalization preprocessing was selected as the first technique for the spectral data with its variables set at 2-norm and length of 1. Thereafter, the orthogonal signal correction (OSC) was selected as against the GLS weighting that was previously used. Then the autoscale replaced the mean centering.

For the measured MC values, autoscale was also used. The reasoning for selecting the OSC is that the GLS weighting gave more weights to the independent variables, which made it cover less of the variance of the dependent variables. This may be considered as a possible reason why the coefficient of determination of prediction was only fair. This reasoning however can only be justified experimentally.

The model was re-calibrated and Table 4.4 shows the variance captured by the recalibrated model.

Table 4.4 Latent variables and their variance captured

No	X-Block LV	X-Block Cumulative	Y-Block LV	y-Block Cumulative	RMSECV Moisture Content	
1	20.0489	20.0489	63.9149	63.9149	0.20479	
2	42.8267	62.8756	15.8027	79.7176	0.18886	
3	11.1524	74.028	11.5443	91.2619	0.1468	
4	6.1232	80.1512	7.5927	98.8546	0.13633	suggested
5	2.5471	82.6983	0.64206	99.4966	0.13831	
6	10.3315	93.0298	0.087875	99.5845	0.13212	current
7	0.65858	93.6884	0.27637	99.8609	0.13047	
8	0.30885	93.9972	0.077799	99.9387	0.1311	
9	0.32758	94.3248	0.030861	99.9696	0.12998	
10	0.37076	94.6955	0.01361	99.9832	0.12978	

It can be seen from Table 4.4 that the software suggested using four LVs for the model. The reasoning for this suggestion is that the RMSECV does not significantly improve when more LVs are selected. However, this suggestion was not accepted because the major reason for recalibrating the model is to capture as much variance as possible within reasonable selection of LVs using the OSC preprocessing. It can be seen that by moving two steps downward in the table, 13% more variance can be captured with a better RMSECV. Hence, this model was calibrated with 6 LVs.

Figure 4.16 shows the RMSEC and RMSECV curve against numbers of LVs.
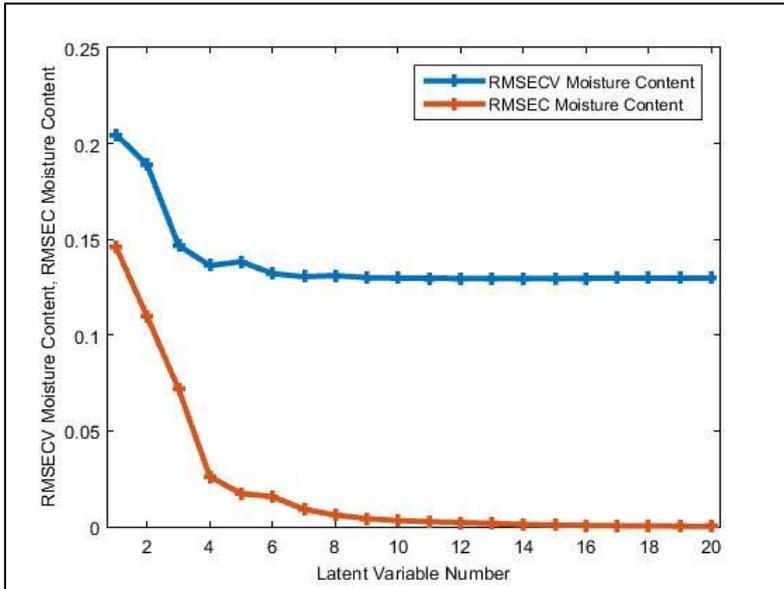
Figure 4.16 RMSECV and RMSEC against LVs

It can be seen from Figure 4.16 that the RMSECV does not significantly improve with more LVs and the improvement in RMSEC is not worth selecting more LVs for. Hence, 6 LVs were retained based on Table 4.4 and Figure 4.16 RMSECV and RMSEC against LVs

The re-calibrated and cross-validated results based on the 6 LVs selected are shown in the figures below:
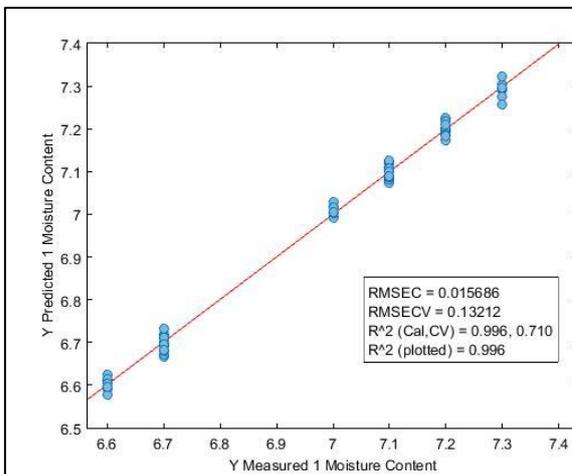
Figure 4.17 Prediction with calibration values   Figure 4.18 Prediction with cross-validation
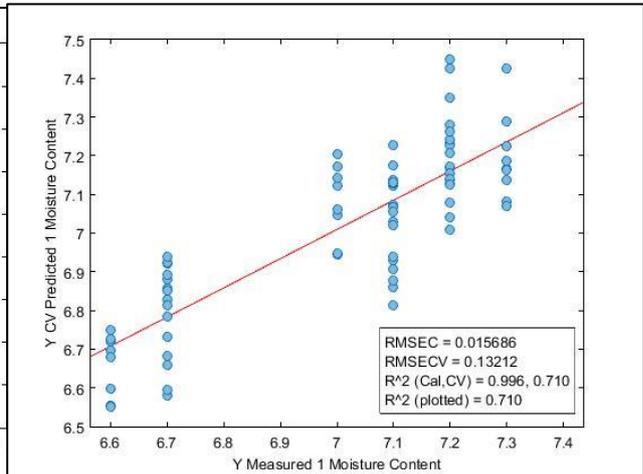
The calibrated model has RMSEC of 0,015686, RMSECV of 0,13212, R^2 of calibration 0,996 and R^2 of cross validation of 0,710. It can be seen that the coefficient of determination of cross validation (R^2) is a little lower than previous calibrations. In this calibration, more of the dependent variable were captured in this model, which is good for the purpose of prediction.

With this model, the test data which were a total of 40 were loaded with their respective MC. After applying the prediction, one of the test data points was an outlier, this was removed and the remaining 39 test data were used to make the prediction. The result is given in the figures below.
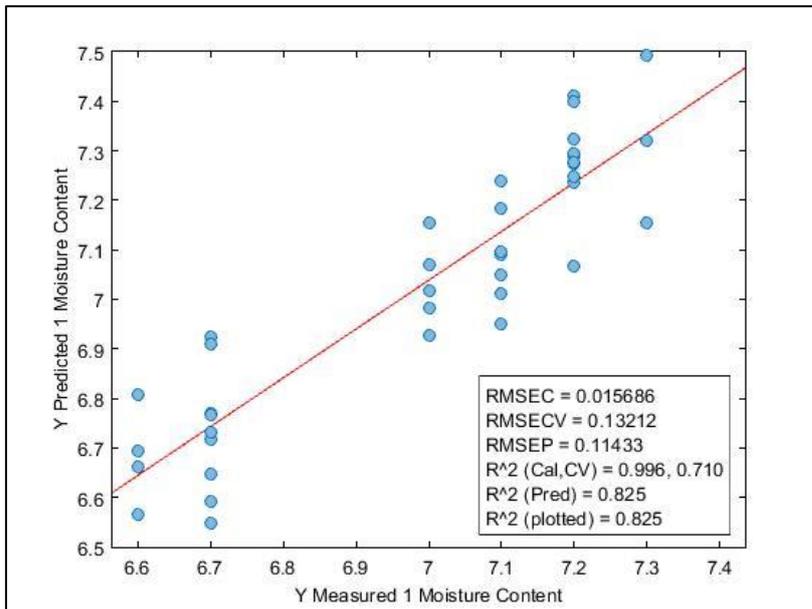
Figure 4.19 Result with the new model

The model predicted the MCs of the spectral data with R^2 of 0,825 and RMSEP OF 0,11433. Additionally, the average of the predicted MCs for all the 39 samples is the same with the average of the measured MC at 7 w-% which is well below the stipulated MC by CEN (< 10 w-%).

It has been explained in literature that the larger the range of wavelengths of spectral data used for the calibration, the higher the R^2 value that can be achieved [6]. Since the aim of this thesis is to achieve results using the range of 420 -900 nm, this can be considered as a good result as the major absorption points of water are beyond these wavelengths. For this thesis, only one subtle absorption point was observed at 540 nm and this was what the regression was based on. Although water peaks were also observed at 620 nm, this was not well defined on all the samples, which did not make it a strong point for distinction.

The predicted data plotted in Figure 4.19 is also given in appendix 3 with the average calculated which is 7 w -% (the same as the measured MC). The result shows little variations around the measured MC, majorly because of the restraints in wavelength used.

In this chapter, the acquired spectral data and measured moisture contents have been used to calibrate a PLSR model which was able to predict moisture content of 39 samples of biofuel pellets with an RMSEP of 0,11433 and R^2 of 0,825.

With this calibrated and validated model, MC of biofuel pellets can be predicted using any set of HSI data with 240 wavelength bands. To do this, the script shown below can be written to matlab with the spectral data and the model already loaded to the workspace.

Pred = pls('my_data','my_model')

Where Pred is the new MC independent results

       Pls calls the function 'pls' to predict 'pred'

       'my data' is the spectral data

       And 'my_model' is the model earlier calibrated [51].

If there is a need however to use a dataset with a different dimension of wavelengths, there has to be another calibrated model. This limitation is indeed peculiar to PLSR as the saved model is in the form where each column is needed for prediction.

5. FUTURE WORKS

This chapter discusses the possibilities for improvement for this work and suggestions of tasks to be undertaken on the robust solution needed for predicting all indices of biofuel pellets using hyperspectral camera. The findings discussed in chapter 4 show a good result of prediction, but the result can be improved upon and a better model can be calibrated. Various ways to achieve a better model will be discussed in this chapter.

The first suggestion to create a better model is to increase the number of types of pellet samples used. The limitation to one type of pellet reduced the ability to regard the result as being a generalized model. A way to make the model more generalized is to include pellets from different sources of raw materials. There was an attempt to use sunflower pellets and somewhat darker wood pellets, but the quantity available was far lesser than the standard 300 g quantity specified by CEN, hence they were not included in the calibration data set.

Another way of improving the model is by using camera with more bands in the NIR region. The fact that only one major absorption point was observed reduced the strength of the calibration and by implication the prediction. Only one out of the twenty five articles reviewed did not use wavelength band of 1000 nm or above. This reduces the accuracy of the model, hence a balance can be found between optimum camera price and efficient wavelength bands for MC prediction. It would have been ideal to search for cameras that take spectral data at just the important wavelengths for MC estimation. However, considering the need to combine other quality indices estimation into the final solution, a more reliable approach would be to select a camera with larger range of wavelengths, more into the NIR region.

To further advance the bigger project, it is worthy of note that HSI has been proposed for other quality indices of biofuel pellets [2]. Gross calorific value (GCV) can be determined by using a similar method to the one used in this thesis. However, in obtaining the reference values to be used for the model calibration, CEN 14918 standard prescription should be followed. Bomb calorimeter, which is calibrated by combusting certified benzoic acid, should be used. The experiment must be done at constant volume with reference temperature of 25°C [14]. This method has been used in literature and the calibrated model had good result [2][22].

In getting the values for ash content, the PLSR model can also be used as it was used in previous research works. Ash content (AC) is the mass of inorganic content remaining in the biofuel pellets after ignition under certain conditions [13]. The method used to obtain the reference value is to burn the pellets under rigidly controlled conditions of time, sample weight and equipment specifications. The temperature should be of (550 ± 10) °C. For this, Gillespie et al. used the furnace OAF 11/1, Carbolite, Hope, United Kingdom [2].

PLSR can also be used to calibrate model for Carbon content (CC). The carbon analyzer (Primacs SLC TOC Analyzer, Model CS22, Skalar Analytical B.V., Breda, The Netherlands) was used in literature to obtain the reference values for CC [2].

The models built in literature for GCV, AC and CC by Gillespie et al. were excellent, good and fair respectively [2].

Apart from these indices, the dimension of the pellets can also be measured using the regular machine vision in the visible wavelength region. The most important part of the task however will be to combine all of these models into a single solution. Importantly, these solutions need to be incorporated into an online monitoring system. A suggestion by the Chair of Mechatronics is to build an application able to use all of the models on a mobile platform. This topic is an exciting future work to improve the solution further.

Another possibility different from a mobile platform is to use a software provided by EigenVector called Solo Predictor[51]. This application is able to acquire data from an online monitoring system and apply a pre-calibrated model to this data to make predictions. Another interesting fact about the software is its ability to apply preprocessing techniques to data if required.

As discussed above, this thesis can be a springboard leading to many other interesting topics, as the strengths and limitations of the work opens the way to other possibilities. Concisely, capabilities for protocol and methodology for hyperspectral imaging in the machine vision laboratory of Electrical Power Engineering and Mechatronics can be improved based on the findings of this thesis work.

6. SUMMARY

Estonia ranks first in the world in the production of biofuel pellets per capita. Hence, the overall objective of this thesis was to consolidate on this achievement and proffer better and faster ways of estimating the quality indices of biofuel pellets. As such, this thesis has examined a novel way of determining moisture content in biofuel pellets. The oven dry method is the prevalent method used in estimating moisture content, however, it is time consuming, destructive and cannot be incorporated into an online production system. Therefore, a review of extant literature helped to establish that hyperspectral imaging technique is the best method to use in estimating the moisture content of biofuel pellets with an incorporated system to an online production system.

Hyperspectral imaging was selected because of its widespread application in various fields such as agriculture to obtain both spatial and spectral information about objects. This gives the ability to acquire information not obtainable in spectroscopy only or in photography. Also, hyperspectral imaging technique has been used to estimate moisture content alongside other parameters in pork, mangoes, potatoes and prawns among others.

In order to estimate moisture content of biofuel pellets, various steps needed to be undertaken. Data for this thesis was gotten from two departments within the university – Electrical Power Engineering and Mechatronics and Energy Technology. In order to get the right estimates, reference data needed to be gotten from department of Energy Technology. The pellet samples were first acquired. This was followed by the setting up of the hyperspectral imaging environment to collect the hyperspectral image data on one hand and the reference data acquisition on the other hand. A wavelength range of 420-900 nm was used in this thesis.

Hyperspectral data were obtained at the Machine Vision laboratory of the Electrical Power Engineering and Mechatronics department. This was done using the Spectronon software and the acquisition camera used was Resonon PIKA II. Ten samples of biofuel pellet spectra were obtained and the average spectra of regions of interest saved for use as calibration data. In a quick succession, the samples were taken to the laboratory of fuel and air emission analysis at the department of energy technology. The moisture contents of the samples were obtained to serve as reference data for analysis. This was done in accordance with the CEN 14774 standard which involves delivering the samples to the laboratory in air tight bags. The pans used for drying were weighed when empty as well as weighed with the samples. After drying for one hour in an atmosphere where the air changes at 105°C for one hour, the samples were weighed again. This was repeated three times every hour until constant mass was achieved for all the samples. The MC was then calculated for all samples.

Following the hyperspectral image acquisition, ROI was selected and then the extraction of raw spectral data was taken. Spectral preprocessing was done to determine outliers to be removed. The initial preprocessing was done in Spectronon pro environment – the same software for the HSI data acquisition. The preprocessing techniques that were used in this task are ROI selection, MSC

(mean), normalization, mean centering, autoscale, Standard normal variate, Savitzky-Golay second derivative filter, GLS weighting and orthogonal signal correction.

The Principal Component Analysis model was developed to critically study the data and ensure that outliers are removed from the data. After the addition of the reference data, the partial least squares regression calibration model was developed. It was then validated and used to predict the estimation of moisture content. The initial calibrated model only achieved fair prediction results, hence, the model was recalibrated to achieve good prediction results. The analysis and prediction results were majorly done with Matlab R2015a using Eigenvector PLS Toolbox.

The acquired spectral data and measured moisture contents were used to calibrate a PLSR model which was able to predict moisture content of 39 samples of biofuel pellets with an RMSEP of 0,11433 and R^2 of 0,825. With this calibrated and validated model, MC of biofuel pellets can be predicted using any set of HSI data with 240 wavelengths. In addition, the average of the predicted MCs for all the 39 samples is the same with the average of the measured MC at 7 w-% which is well below the stipulated MC by CEN (< 10 w- %). This is a good result because the aim of this thesis is to achieve results using the range of 420-900 nm and the major absorption points of water are beyond these wavelengths. In this thesis, only one subtle absorption point was observed at 540 nm. This was what the PLSR was based on. Although water peaks were also observed at 620 nm, this was not well defined on all the samples, which did not make it a strong point for distinction.

The goal of this thesis was to build a model using the visible and near infrared wavelengths. Therefore, the findings of this thesis make significant contributions to the bigger project in the Department of Electrical Power Engineering and Mechatronics to create a machine vision system able to estimate the quality indices of biofuel pellets in one exposure during the production routine.

7. KOKKUVÕTE

Eesti hoiab maailma esikohta biokütuse graanulite tootmises ühe elaniku kohta. Selle lõputöö eesmärk on pakkuda paremat ja kiiremat viisi kraanulite kvaliteedi tõstmiseks. Selleks on käesolevas lõputöös uuritud niiskuse mõõtmise viise biokütuse pelletites. Ahjus kuivatamise meetod on küll levinud, kuid on ajaklukas, hävitav ning seda pole võimalik tootmisliinis tootmisprotsesside ajal kasutada. Olemasoleva kirjanduse ülevaade näitab, et hüperspektraalset tehnikat on potentsiaalne viis, et määrata niiskusesisaldust biokütuse.

Valituks sai hüperspektraalne pilditehnika tänu selle laialdasele rakendusele erinevates valdkondades. Näiteks kasutatakse seda põllumajanduses, et saada erinevatest objektidest nii ruumilist, kui ka spektraalset informatsiooni. Nimetatud viis võimaldab omandada infot mitte ainult nähtavas spektroskoopias, aga ka fototehnikas. Lisaks kasutatakse hüperspektraalset tehnikat niiskuse ning teiste parameetrite hindamiseks ka toiduainete puhul nagu näiteks sealiha, mangod, kartulid, krevetid jt.

Niiskussisalduse hindamiseks biokütuse pelletites vajab mitmeid eelnevaid samme. Andmed käesoleva diplomitöö koostamiseks on hangitud kahest ülikooli osakonnast – elektrotehnika ja mehhatroonika instituudist aga ka energeetika instituudist. Selleks, et saada õiged andmed, oli vaja saada võrdlusandmed energiatehnoloogia intstituudi teadurite abiga. Esiteks oli vaja pelletite proove. Sellele järgnes hüperspektraalse hõive jaoks keskkonna loomine, hüperspektraalse pildi loomiseks ning võrdlusandmete saamiseks. Mõõtmiste lainepikkuste vahemik oli 420-900 nm.

Hüperspektraalsed andmed saadi elektroenergeetika ja mehhatroonika instituudi masinnägemise laboris. Selleks kasutati Spectronon tarkvara ja Resonon PIKA II kaamerat. Koguti kümme biokütuse pelleti spektrit ning vajalike piirkondade keskmist kasutati kalibreerimisandmetena. Järgnevalt viidi proovid kütuse ja õhu emissiooni analüüsi teaduslaborisse energiatehnoloogia instituudis. Niiskuskeskkond püüti säilitada võimalikult sarnane eesmärgiga talletada võrdlusmaterial. Seda tehti vastavuses standardiga CEN 14774 õhukindlate kottide abil. Võeti ka kuivatamisel kasutatavate pannide tühikaal enne objektide kaalumist. Peale tunnist 105 °C juures kuivatamist kaaluti graanuleid uuesti. Viimast korrati kolm korda iga tunni jooksul kuni mass graanulite kaalumisel enam ei muutunud. Iga proovi jaoks koostati niiskussisalduse arvutus.

Peale hüperspektraalse pildi võtmist valiti ROI, mille järel oli võimalik eraldada vajalikud spektriandmed. Ebavajalike andmete eemaldamiseks kasutati eeltöötlust. Esialgne töötlus tehti Spectronon Pro keskkonnas. Sama tarkvara kasutati ka HSI andmete puhul. Vajalikud funktsionaalsused olid ROI valik, MSC, normaliseerimine, keskmise tsentreerimine, automaatne skaleerimine, standardi kaudu normaliseerimine, Savitzky-Golay teise tuletise filter, GLS kaalumine ja ortogonaalse signaali korrektsioon.

Loodi põhikomponentide analüüsil (PCA) põhinev mudel, et uurida andmeid ja tagada valeandmete eemaldamine. Peale võrdlusandmete lisamist loodi osalise vähimruutude regressiooni (PLS) kalibratsiooni mudel. Seejärel see valideeriti ja kasutati niiskuse sisalduse ennustamisel. Esialgsed

tulemused ei olnud piisavalt täpsed, mistõttu oli vajalik mudeli uuesti kalibreerimine. Viimane parandas ennustustäpsust. Analüüsid ja prognoosid olid põhiliselt Matlab R2015a keskonnas kasutades Eigenvector PLS ToolBoxi.

Omandatud spektriandmeid ja niiskusesisalduse mõõtmisi kasutati, et kalibreerida PLSR mudel, mis võimaldas ennustada niiskussisaldust 39 biokütuse pelletite proovis, kus RMSEP = 0,11433 ja R2 = 0,825. Selle kalibreeritud ja kontrollitud mudeliga võib prognoosida niiskussisaldust biokütuse pelletites kasutades mistahes HSI andmeid sarnastel lainepikkustel. Ennustatud niiskusesisaldus kõigil 39 proovil on keskmiselt 7 massiprotenti. See on sama kui keskmine mõõdetud niiskusesisaldus ja see tulemus on tunduvalt väiksem kui nõutud väärtust CEN, mis on <10 massiprotsenti. See on hea tulemus, sest selle töö eesmärk on saavutada arvestatavad tulemused kasutades 420-900 nm piirkonda. Enamus vee spektaalseid neeldumispunkte on sellest piirkonnast väljaspool. Selles töös uuriti 540 nm piirkonnas vaid üht vee neeldumispunkti. See oli see, millel PLSR põhines. Vee neeldumispunktid esinesid ka 620nm pikkusel, kuid seda ei määratletud kõikides proovides, seega jäeti see suurema tähelepanuta.

Selle töö eesmärk oli ehitada mudel kasutades nähtavaid ja lähisinfrapunaseid lainepikkusi biokütuse graanulite niiskuse sisalduse määramiseks. Töö järeldusi saab kasutada elektroenergeetika ja mehhatroonika instituudi mahukamas projektis, milel eesmärgiks luua masinnägemine süsteem, mis võimaldab hinnata biokütuse pelletite kvaliteeti jooksvalt graanulite tootmise käigus.

8. REFERENCES

[1]   E. Alakangas, E. Iii, and C. E. N. Tc, "New European Pellets Standard – EN 14961-1," vol. 5, pp. 1–9.

[2]   G. D. Gillespie, C. D. Everard, and K. P. Mcdonnell, "Prediction of biomass pellet quality indices using near infrared spectroscopy," *Energy*, vol. 80, pp. 582–588, 2015.

[3]   M. Guo, W. Song, and J. Buhain, "Bioenergy and biofuels : History , status , and perspective," vol. 42, pp. 712–725, 2015.

[4]   K. Hermann and R. Dietmar, *A history of India*, 5th ed. New York: NY: Routledge, 2010.

[5]   P. Sheperd, *Biomass co-firing: A renewable alternative for utilities*. Golden, Colorado, 2000.

[6]   J. P. Carroll and J. Finnan, "Physical and chemical properties of pellets from energy crops and cereal straws," *Biosyst. Eng.*, vol. 112, no. 2, pp. 151–159, 2012.

[7]   M. Arshadi, R. Gref, P. Geladi, S. Dahlqvist, and T. Lestander, "The influence of raw material characteristics on the industrial pelletizing process and pellet quality," *Fuel Process. Technol.*, vol. 89, no. 12, pp. 1442–1447, 2008.

[8]   E. Standard, "EUROPÄISCHE NORM FINAL DRAFT FprEN 14774-2 Solid biofuels - Determination of moisture content - Oven dry method - Part 2 : Total moisture - Simplified method," 2009.

[9]   M. Kamruzzaman, Y. Makino, and S. Oshita, "Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging," *Food Chem.*, vol. 196, pp. 1084–1091, 2016.

[10]   H. Kobori, N. Gorretta, G. Rabatel, V. Bellon-maurel, and G. Chaix, "Applicability of Vis-NIR hyperspectral imaging for monitoring wood moisture content (MC)," vol. 67, no. 3, pp. 307–314, 2013.

[11]   S. R. Sandmeier and K. I. Itten, "A field goniometer system (FIGOS) for acquisition of hyperspectral BRDF data," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2 II, pp. 978–986, 1999.

[12]   S. Hinterreiter, H. Hartmann, and P. Turowski, "Bridging properties of biomass fuels," *Energy Prod.*, pp. 280–282, 2010.

[13]   E. Alakangas, "European standards for solid biofuels Fuel specification and classes , multipart standard," 2008.

[14]   SolidStandards, "Enhancing the implementation of quality and sustainability standards and certification schemes for solid biofuels (EIE/11/218)," 2011.

[15]   H. Huang, L. Liu, and M. O. Ngadi, "Recent Developments in Hyperspectral Imaging for

Assessment of Food Quality and Safety," pp. 7248–7276, 2014.

[16]   D. Bannon, "cubes and slices," *Nat. Photonics*, vol. 3, no. 11, pp. 627–629, 2009.

[17]   L. Gao and R. T. Smith, "Optical hyperspectral imaging in microscopy and spectroscopy - a review of data acquisition," *J. Biophotonics*, vol. 456, no. 6, pp. 441–456, 2014.

[18]   W. Cho, "Hyperspectral Data Acquisition and Its Application for Face Recognition," 2015.

[19]   E. Villeneuve, H. Carfantan, and D. Serre, "PSF estimation of hyperspectral data acquisition system for ground-based astrophysical observations," *Work. Hyperspectral Image Signal Process. Evol. Remote Sens.*, 2011.

[20]   C. Rogass, C. Mielke, D. Scheffler, N. K. Boesche, A. Lausch, C. Lubitz, M. Brell, D. Spengler, A. Eisele, K. Segl, and L. Guanter, "Reduction of uncorrelated striping noise-applications for hyperspectral pushbroom acquisitions," *Remote Sens.*, vol. 6, no. 11, pp. 11082–11106, 2014.

[21]   S. Deng, Y. Xu, X. Li, and Y. He, "Moisture content prediction in tealeaf with near infrared hyperspectral imaging," *Comput. Electron. Agric.*, vol. 118, pp. 38–46, 2015.

[22]   C. C. Fagan, C. D. Everard, and K. Mcdonnell, "Prediction of moisture , calorific value , ash and carbon content of two dedicated bioenergy crops using near-infrared spectroscopy," *Bioresour. Technol.*, vol. 102, no. 8, pp. 5200–5206, 2011.

[23]   C. Rhe, "Multivariate NIR spectroscopy models for moisture , ash and calorific content in biofuels using bi-orthogonal partial least squares regression," pp. 1182–1189, 2005.

[24]   T. A. Lestander, B. Johnsson, and M. Grothage, "NIR techniques create added values for the pellet and biofuel industry," *Bioresour. Technol.*, vol. 100, no. 4, pp. 1589–1594, 2009.

[25]   M. Zhao, C. Esquerre, G. Downey, and C. P. O. Donnell, "Process analytical technologies for fat and moisture determination in ground beef - a comparison of guided microwave spectroscopy and near infrared hyperspectral imaging," *Food Control*, vol. 73, pp. 1082–1094, 2017.

[26]   J. Qu, D. Sun, J. Cheng, and H. Pu, "Mapping moisture contents in grass carp (Ctenopharyngodon idella) slices under different freeze drying periods by Vis-NIR hyperspectral imaging," *LWT - Food Sci. Technol.*, vol. 75, pp. 529–536, 2017.

[27]   Y. Pu and D. Sun, "Combined hot-air and microwave-vacuum drying for improving drying uniformity of mango slices based on hyperspectral imaging visualisation of moisture content distribution," *Biosyst. Eng.*, vol. 156, pp. 108–119, 2017.

[28]   J. Ma, D. Sun, and H. Pu, "Model improvement for predicting moisture content (MC) in pork longissimus dorsi muscles under diverse processing conditions by hyperspectral imaging," vol. 196, pp. 65–72, 2017.

[29]   Y. Zhu, X. Zou, T. Shen, J. Shi, J. Zhao, M. Holmes, and G. Li, "Determination of total acid content and moisture content during solid-state fermentation processes using hyperspectral imaging," *J. Food Eng.*, vol. 174, pp. 75–84, 2016.

[30]   W. Su and D. Sun, "Potential of hyperspectral imaging for visual authentication of sliced organic potatoes from potato and sweet potato tubers and rapid grading of the tubers according to moisture proportion," *Comput. Electron. Agric.*, vol. 125, pp. 113–124, 2016.

[31]   Y. Pu and D. Sun, "Prediction of moisture content uniformity of microwave-vacuum dried mangoes as affected by different shapes using NIR hyperspectral imaging," vol. 33, pp. 348–356, 2016.

[32]   T. Möckel, O. Löfgren, H. C. Prentice, L. Eklundh, and K. Hall, "Airborne hyperspectral data predict Ellenberg indicator values for nutrient and moisture availability in dry grazed grasslands within a local agricultural landscape," *Ecol. Indic.*, vol. 66, pp. 503–516, 2016.

[33]   J. Ma, D. Sun, and H. Pu, "Spectral absorption index in hyperspectral image analysis for predicting moisture contents in pork longissimus dorsi muscles," vol. 197, pp. 848–854, 2016.

[34]   Y. Yang, D. Sun, and N. Wang, "Rapid detection of browning levels of lychee pericarp as affected by moisture contents using hyperspectral imaging," vol. 113, pp. 203–212, 2015.

[35]   Y. Pu and D. Sun, "Vis – NIR hyperspectral imaging in visualizing moisture distribution of mango slices during microwave-vacuum drying," vol. 188, pp. 271–278, 2015.

[36]   D. M. M. Gila, P. C. Marchal, J. G. García, and J. G. Ortega, "On-line system based on hyperspectral information to estimate acidity , moisture and peroxides in olive oil samples," vol. 116, pp. 1–7, 2015.

[37]   C. Garrido-novell, A. Garrido-varo, D. Pérez-marín, J. E. Guerrero-ginel, and M. Kim, "Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging," vol. 153, pp. 117–123, 2015.

[38]   D. Liu, D. Sun, J. Qu, X. Zeng, H. Pu, and J. Ma, "Feasibility of using hyperspectral imaging to predict moisture content of porcine meat during salting process," vol. 152, pp. 197–204, 2014.

[39]   M. Huang, Q. Wang, M. Zhang, and Q. Zhu, "Prediction of color and moisture content for vegetable soybean during drying using hyperspectral imaging technology," vol. 128, pp. 24–30, 2014.

[40]   A. Iqbal, D. Sun, and P. Allen, "Prediction of moisture , color and pH in cooked , pre-sliced turkey hams by NIR hyperspectral imaging system," vol. 117, pp. 42–51, 2013.

[41]   H. He, D. Wu, and D. Sun, "Non-destructive and rapid analysis of moisture distribution in farmed Atlantic salmon (Salmo salar) fi llets using visible and near-infrared hyperspectral imaging," vol. 18, pp. 237–245, 2013.

[42]   D. Wu, H. Shi, S. Wang, Y. He, Y. Bao, and K. Liu, "Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system," vol. 726, pp. 57–66, 2012.

[43]   Q. Wang, P. Li, Z. Pu, and X. Chen, "Calibration and validation of salt-resistant hyperspectral indices for estimating soil moisture in arid land," vol. 408, pp. 276–285, 2011.

[44]   J. L. Marshall, P. Williams, J. Rheault, T. Prochaska, R. D. Allen, and D. L. Depoy, "Characterization of the Reflectivity of Various Black Materials," vol. 9147, pp. 1–8, 2014.

[45]   M. Sillat, "APPLICATION OF HYPERSPECTRAL IMAGING IN WASTE PAPER SORTING," 2016.

[46]   Riley Surface World, "Gallenkamp (Sanyo / Weiss) Hot Box Size 2 Oven The Marketplace for Surface Technology . Gallenkamp (Sanyo / Weiss) Hot Box Size 2 Oven." [Online]. Available: http://www.rileysurfaceworld.co.uk/live/machines2/23653.pdf. [Accessed: 05-May-2017].

[47]   F. Van Den Berg and S. B. Engelsen, "Review of the Most Common pre-Processing Techniques for Near-Infrared Spectra Review of the most common pre-processing techniques for near-infrared spectra," *Trends Anal. Chem.*, vol. 28, no. 10, pp. 1201–1222, 2009.

[48]   V. Bochko, "Preprocessing : Smoothing and Derivatives," 2010.

[49]   B. M. Wise, N. B. Gallagher, and W. Windig, *Chemometrics Tutorial for PLS _ Toolbox and Solo*. .

[50]   D. A. Burns and E. W. Ciurczak, *Handbook of near-infrared analysis*, 3rd ed. CRC press, 2007.

[51]   "Eigenvector PLS_Toolbox for MatLab." [Online]. Available: http://eigenvector.com/. [Accessed: 08-May-2017].

[52]   D. K. Sarma, M. Konwar, J. Das, S. Pal, and S. Sharma, "A Soft Computing Approach for Rainfall Retrieval From the TRMM Microwave Imager," vol. 43, no. 12, pp. 2879–2885, 2005.

[53]   B. Mevik and R. Wehrens, "Introduction to the pls Package," no. Section 7, 2016.

9. APPENDICES

Appendix 1. Resonon Pika II datasheet

Pika II (400 – 900 nm)

The **Pika II** is our most popular and most affordable hyperspectral imaging camera.

Pika II Specifications

Spectral Range	400 – 900 nm
Spectral Resolution *	2.1 nm
Spectral Channels	240
Spatial Channels	640
Max Frame Rate	145 fps
Bit Depth	12
Weight	2.8 lbs, 1.3 kg
Dimensions	9.7 x 16.8 x 6.4 cm
Connection Options	GigE
Temperature Range	46-90 F, 8-32 C
f/#	f / 3.0
Avg. RMS Spot Radius	7 µm
Smile (peak-to-peak)	5 µm
Keystone (peak-to-peak)	7 µm

*The number of spectral channels equals the spectral range divided by the spectral resolution, and depends on the RMS spot size. The number of independent spectral channels is NOT the same as the number of sensor pixels in the spectral direction.
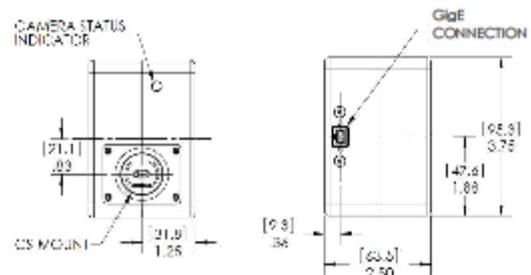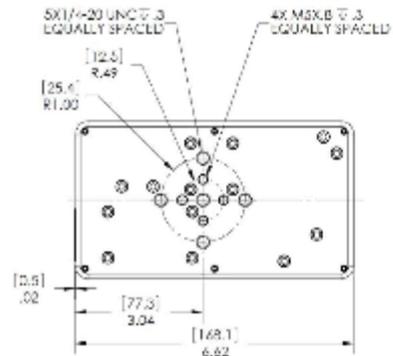
www.resonon.com   inquiry@resonon.com   1.406.586.3356

Appendix 2. Xenoplan 1.4/23 datasheet

Schneider KREUZNACH

Ruggedized Lens

Xenoplan 1.4/23 - Ruggedized

In accordance with the sensitivity of modern 2 / 3" CCD and CMOS sensors, the 3 megapixel lenses are corrected and broadband-coated for the spectral range of 400 – 1000 nm (VIS + NIR). Even under production and / or extreme conditions, the robust mechanical design with lockable focus and iris setting mechanism guarantees reliable continuous use in which the set optical parameters remain in place.

Xenoplan 1.4/23

Key Features

- High-resolution optics
- Stabilized mechanism
- Highest optical imaging performance even with smallest pixel sizes
- Broadband coating (400 - 1000 nm)
- Compact and low weight
- Vibration insensitivity for stable imaging performance, secured ring
- Focus and iris setting lockable

Applications

- 3D measurement
- Machine Vision and other imaging applications
- Traffic
- Medical
- Robot vision
- Food processing

Technical Specifications

F-number	1.4
Focal length	22.5 mm
Image circle	11 mm
Transmission	400 - 1000 nm
Interface	C-Mount
Weight	94 gr.
Option	Optical filter

Appendix 3. Measured MC and Predicted MC

Table 9.1 Measured and predicted MC

No	Measured MC	Predicted MC
1	7.2	7.4
2	7.2	7.3
3	7.2	7.3
4	7.2	7.4
5	7.2	7.1
6	7.2	7.2
7	7.2	7.2
8	7.2	7.3
9	7.2	7.3
10	7.3	7.3
11	7.3	7.5
12	7.3	7.2
13	7	6.9
14	7	7.2
15	7	7.0
16	7	7.0
17	7	7.1
18	7.1	7.1
19	7.1	7.0
20	7.1	7.1
21	7.1	7.2
22	6.7	6.5
23	6.7	6.9
24	6.7	6.8
25	6.7	6.9
26	7.1	7.1
27	7.1	7.2
28	7.1	7.0
29	7.1	7.0
30	7.2	7.3
31	6.7	6.8
32	6.7	6.6
33	6.7	6.6
34	6.7	6.7
35	6.7	6.7
36	6.6	6.7
37	6.6	6.7
38	6.6	6.6
39	6.6	6.8
Average	**7.0**	**7.0**

Appendix 4. Field of view used
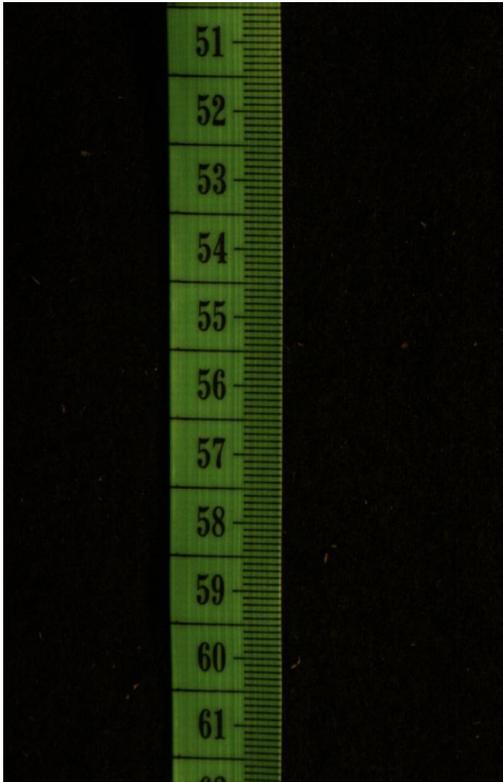
Figure 9.1 Field of view Y direction
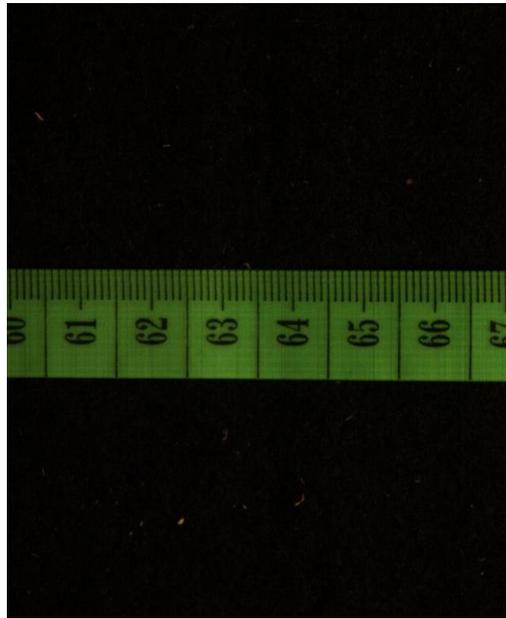
Figure 9.2 Field of view X direction

Appendix 5. Variance for GLS weighted calibrated model

Table 9.2 Variance Table after applying GLS weighting

No	X-Block LV	X-Block Cumulative	Y-Block LV	y-Block Cumulative	RMSECV Moisture Content	
1	10.548	10.548	83.6368	83.6368	0.15021	
2	3.1926	13.7406	11.1345	94.7713	0.15098	
3	2.0014	15.742	3.8133	98.5845	0.14062	
4	1.936	17.678	0.93924	99.5238	0.13442	
5	1.6704	19.3485	0.33785	99.8616	0.13128	current*
6	1.8271	21.1756	0.08566	99.9473	0.1318	
7	1.9408	23.1163	0.034149	99.9814	0.13109	
8	1.9196	25.0359	0.011644	99.9931	0.13081	
9	1.5275	26.5634	0.0047691	99.9978	0.13062	
10	2.0013	28.5647	0.0011518	99.999	0.13061	

Appendix 6. Technical parameters of the pellets used

Figure 9.3 Technical parameters of pellets used

Appendix 7. Description of Included files

A folder with the name of the thesis is submitted together with this work. The following files are contained in this folder:

1. A folder named 'hypercubes' containing the obtained hypercubes for all ten classes, provided as .bil and .hdr files are provided. The RGB images are also added.
2. A folder named 'Mean spectra' containing the mean spectra data used for both calibration and validation are provided in .xls files.
3. A folder named 'matlab files' containing the calibration, validation datasets and preprocessing techniques used in .mat format as matlab files.
4. The final PLSR model built, in .mat format
5. A folder "Articles" containing all scientific papers used in the thesis.