Università Politecnica delle Marche

Ph.D. in Agriculture, Food and Environmental Sciences

XXX cycle

Application of infrared spectroscopy technique for the assessment of solid biofuels quality

Ph.D. Thesis

Manuela Mancini

Tutor:

Prof. Giuseppe Toscano

Co-tutor:

Prof. Åsmund Rinnan

November 2017

Preface

This Ph.D. thesis, titled ‘Application of infrared spectroscopy technique for the assessment of solid biofuels quality’, is submitted as a requirement for obtaining the Ph.D. Degree from the Ph.D. School of Agriculture, Food and Environmental Sciences (Università Politecnica delle Marche), with the further mention of *Doctor Europaeus*.

The work presented in this Ph.D. thesis has been carried out during the three-years research activities conducted by the author, Manuela Mancini, at the Department of Agriculture, Food and Environmental Sciences, Università Politecnica delle Marche (Ancona), under the supervision of Prof. Giuseppe Toscano. Part of the work has been done at the Department of Food Science (Chemometrics and Analytical Technology Section), University of Copenhagen, under the supervision of Prof. Åsmund Rinnan.

Manuela Mancini
Ancona, November 2017

Acknowledgements

First of all, I would like to give a special thank to all the research group. In particular, it is a pleasure to thank Prof. Giuseppe Toscano who trusted me and gave me the opportunity to perform these three years research. He encouraged me in trying and working, always showing great enthusiasm and strongly believing in the successful of our research. I am really grateful to all the Biomass Lab group who teaches me not only about science and laboratory analysis, but also the priceless value of a good cooperation between colleagues. Thank you for all the laughs during the lunch time, for all the suggestions and for always giving me help.

I would like to give a huge "Tak" to Prof. Åsmund Rinnan who has been my supervisor during my six-months stay at the Chemometrics & Analytical Technology section (CAT) of the Department of Food Science (University of Copenhagen). He has introduced me to the marvellous, crazy world of chemometrics and I am very much grateful not only for the pleasure to work together, but also for going on with the collaboration and for being always carried away with new research ideas. I would like to thank also all the CAT section for the nice moments and for all the friends and nice people that gave me the opportunity to meet. I never expected such a friendly and welcoming atmosphere well mixed with high quality science.

These acknowledgment will be not complete without mentioning my friends. Some are close also by distance, others just by mind. In any case, all of them are just so wonderful in a thousand of different ways and I can not ask for more than that.

Finally, my deepest acknowledgment belongs to my family. Absolutely not normal, sometimes over the top, but I would not change for anything. Thank you for always being there, for making me laugh and keep my thoughts out away. I dedicate this thesis to my grandparents, that even if they are no more with me, they still are my roots and my force.

List of publications

Paper I

Duca, D., M. Mancini, G. Rossini, C. Mengarelli, E. Foppa Pedretti, G. Toscano and A. Pizzi (2016). "Soft Independent Modelling of Class Analogy applied to infrared spectroscopy for rapid discrimination between hardwood and softwood." *Energy* **117**: 251-258.

Paper II

Duca, D., A. Pizzi, G. Rossini, C. Mengarelli, E. Foppa Pedretti and M. Mancini (2016). "Prediction of Hardwood and Softwood Contents in Blends of Wood Powders Using Mid-Infrared Spectroscopy." *Energy & Fuels* 30(4): 3038-3044.

Paper III

Toscano, G., Å. Rinnan, A. Pizzi and M. Mancini (2017). "The Use of Near-Infrared (NIR) Spectroscopy and Principal Component Analysis (PCA) To Discriminate Bark and Wood of the Most Common Species of the Pellet Sector." *Energy & Fuels* 31(3): 2814-2821.

Paper IV

Mancini, M., Å. Rinnan, A. Pizzi and G. Toscano (2018). "Prediction of gross calorific value and ash content of woodchip samples by means of FT-NIR spectroscopy." *Fuel Processing Technology* 169: 77-83.

Paper V

Mancini, M., Å. Rinnan, A. Pizzi and G. Toscano, "Use Of FT-NIR Spectroscopy For The Detection Of Residues From Wood Processing Industry In The Pellet Sector". 2017. *Conference proceedings submitted* in: ICNIRS 2017 DENMARK – Copenhagen 11-15 June 2017.

Paper VI

Mancini, M., Å. Rinnan, A. Pizzi, C. Mengarelli, G. Rossini, D. Duca and G. Toscano (2017). "Near infrared spectroscopy for the discrimination between different residues of the wood processing industry in the pellet sector". *Fuel. Manuscript under review.*

Paper VII

Mancini, M., G. Toscano and Å. Rinnan. "Study of the scattering effects on NIR data for the prediction of ash content using MSC correction factors". *Manuscript draft.*

Other publications

Toscano, G., D. Duca, E. Foppa Pedretti, A. Pizzi, G. Rossini, C. Mengarelli and M. Mancini (2016). "Investigation of woodchip quality: Relationship between the most important chemical and physical parameters." *Energy* **106**: 38-44.

Duca, D., A. Pizzi, M. Mancini, G. Rossini, C. Mengarelli, A. Ilari, G. Lucesoli, G. Toscano and E. Foppa Pedretti (2016). "Fast measurement by infrared spectroscopy as support to woody biofuels quality determination." *Journal of Agricultural Engineering* **47**(1): 5.

Pizzi, A., G. Toscano, E. Foppa Pedretti, D. Duca, G. Rossini, C. Mengarelli, A. Ilari, A. Renzi and M. Mancini (2017). "Energy characteristics assessment of olive pomace by means of FTNIR spectroscopy." *Energy. Manuscript under review.*

Pizzi, A., E. Foppa Pedretti, D. Duca, G. Rossini, C. Mengarelli, A. Ilari, M. Mancini and G. Toscano (2017). "Emissions of heating appliances fuelled with agropellet produced from vine pruning residues and environmental aspects." *Renewable Energy. Manuscript under review.*

Foppa Pedretti, E., D. Duca, G. Toscano, C. Mengarelli, G. Rossini, A. Pizzi, M. Mancini, D. Tesei and A. Ilari (2017). "Validity of the mechanical threshing of onion seeds from the point of view of seed quality." *Agriculture. Manuscript accepted for publication.*

Conference proceedings: M. Mancini, A. Pizzi, Å. Rinnan, C. Mengarelli, D. Duca, G. Rossini, E. Foppa Pedretti, A. Ilari and G. Toscano (2016). "Spettroscopia nel vicino infrarosso e analisi multivariata dei dati come supporto per la determinazione della qualità del pellet". In: Proc. NIR ITALIA 2016 ATTI DEL SIMPOSIO. 7 Simposio italiano di Spettroscopia NIR. Milano, Ottobre 2016. ISBN 978-88-941-1531-4.

Abstract in conference proceedings: Mancini M., A. Pizzi, C. Mengarelli, G. Rossini, E. Foppa Pedretti, G. Toscano, A. Renzi and D. Duca (2016). "Energy characteristics assessment of residues by means of infrared spectroscopy." In: Mechanization and new technologies for the control and sustainability of agricultural and forestry systems MECHTECH 2016 Conference, Alghero, Italy, 29th May-1st June 2016. ISBN 979-12-200-1098-6.

Author’s contribution

Paper I: Dr. Andrea Pizzi and Dr. Daniele Duca had the main responsibility for planning the work, interpretation of the results and writing the publication under the supervision of Prof. Giuseppe Toscano and Prof. Ester Foppa Pedretti. The author performed the laboratory work and the infrared analysis with the help of Dr. Chiara Mengarelli, Dr. Giorgio Rossini and Dr. Andrea Pizzi. The author assisted and helped in the multivariate data analysis and in the writing process.

Paper II: the author had the main responsibility for planning the work, interpretation of the results and writing the publication under the supervision of Prof. Ester Foppa Pedretti. The author performed the laboratory work and infrared analysis with the help of Dr. Chiara Mengarelli, Dr. Giorgio Rossini and Dr. Andrea Pizzi. The author with the help of Dr. Andrea Pizzi had the main responsibility for the multivariate data analysis and infrared analysis. Dr. Duca assisted and helped in the organization of the experimental work and in the writing process.

Paper III: the author was responsible for planning the work, interpretation of the results and writing the publication under the supervision of Prof. Åsmund Rinnan and Prof. Giuseppe Toscano. The laboratory work and infrared analysis were performed by the author and Dr. Andrea Pizzi. The author and Prof. Åsmund Rinnan had the main responsibility for the multivariate data analysis.

Paper IV: the author had the main responsibility for planning the work, interpretation of the results and writing the publication together with Prof. Åsmund Rinnan. The author carried out the laboratory analysis and infrared analysis together with Dr. Andrea Pizzi and under the supervision of Prof. Giuseppe Toscano.

Paper V: the author had the main responsibility for planning the work, interpretation of the results and writing the publication under the supervision of Prof. Åsmund Rinnan and Prof. Giuseppe Toscano. The laboratory work and infrared analysis were performed by the author and Dr. Andrea Pizzi. The author with the help of Prof. Åsmund Rinnan had the main responsibility for the multivariate data analysis.

Paper VI: the author was responsible for planning the work, interpretation of the results and writing the publication under the supervision of Prof. Åsmund Rinnan and Prof. Giuseppe Toscano. The author performed the laboratory work and the infrared analysis with the help of Dr. Andrea Pizzi. Prof. Åsmund Rinnan assisted in the modelling studies. Dr. Chiara Mengarelli and Dr. Giorgio Rossini assisted and helped in the organization of the experimental work and Dr. Daniele Duca in the writing process.

Paper VII: the author was responsible for planning the work, interpretation of the results and writing the publication together with Prof. Åsmund Rinnan. The author and Prof. Åsmund Rinnan had the main responsibility for the multivariate data analysis and infrared analysis. The work was carried out under the supervision of Prof. Giuseppe Toscano.

List of abbreviations

Ac	Ash content
EMSC	Extended Multiplicative Signal Correction
FT	Fourier transform
GCV	Gross calorific value
iPLS	interval Partial Least-Squares
IR	Infrared
LV	Latent variable
Mc	Moisture content
MIR	Medium infrared
MVA	Multivariate data analysis
NIR	Near infrared
OT	Overtone
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares regression
PLS-DA	Partial Least Squares - Discriminant Analysis
Q^2	Coefficient of determination (calculated on predicted / cross-validated values)
R^2	Coefficient of determination (calculated on calibration values)
RMSE	Root Mean Square Error (calculated on calibration values)
RMSECV	Root Mean Square Error of Cross Validation
RMSEP	Root Mean Square Error of Prediction

ROC	Receiver Operating Characteristic
rPLS	recursive weighted Partial Least Squares regression
SIMCA	Soft Independent Modelling of Class Analogy
VIP	Variables Important in Projections

Abstract

Energy production is one of the main important causes of greenhouse gases emission and it can be a huge problem considering the continuous worldwide increase in energy demand. Aware of the importance of finding new ways of clean energy production and of the necessity to limit the dependency on fossil fuels, European policy currently promotes renewable energy sources. Among the different types, solid biofuels - in the form of pellet, woodchip, briquettes and others – are some of the most promising alternative.

The fast growth of the biofuels market highlights the problem of product quality, which has strong repercussions for technical, environmental, and economic aspects. Moreover, the variability of the material characteristics leads to the need to check its quality in order to ensure a good product and make it competitive with the other fuels. In this context, infrared spectroscopy aims to be a new method of quality control. Considering the economic returns of the sector, conventional laboratory analyses are simply too costly. Furthermore, they are long, complex and require skilled operator, not matching the real necessities of the market that needs to have the results as fast as possible.

This Ph.D. project focused on developing an alternative low cost method of biomass quality control based on infrared spectroscopy. Indeed, the first part of the thesis includes four studies of application of infrared spectroscopy to investigate the qualitative parameters of biofuels. In particular, the first and second studies concern the discrimination between hardwood and softwood samples and the subsequent prediction of hardwood/softwood content in blend samples. The third and fourth studies are related to the discrimination between virgin wood and residues from wood processing industries, i.e. bark and treated wood.

The second part of the project shows examples of application of near infrared spectroscopy for the prediction of the quantitative parameters of biofuels, in particular ash content, gross calorific value and moisture content. The studies were performed using both a laboratory spectrophotometer and an on-line instrument customized for laboratory analysis. A deepening was carried out on the influence of scattering effect for the ash content prediction.

The results are preliminary, but really promising. The outcomes of this Ph.D. project confirm the possibility to use infrared spectroscopy, coupled with chemometrics, as a screening tool for getting information about the main parameters of biomass, ensuring a more effective and extensive monitoring of the quality control, especially in the crucial steps of the bioenergy chain.

Table of contents

1. Introduction

1.1 Background and aim

Currently, biomass is the third largest primary energy resource in the world after coal and oil (Purohit, Tripathi et al. 2006) and can include different types of material: plant forest, residues from agricultural and forest production and organic waste from industry, humans and animals. Biomass contributed approximately 5.7% to the European Union's energy consumption in 2008 (European Union 2009), predominately through heat and power applications. Considering the concern about greenhouse gas emissions, the increasing energy demand, the decreasing availability of fossil fuels and the variability of their price, European public policy promotes the energy production from renewable resources. One of the targets is that 20% for the share of energy consumption has to come from renewable sources by 2020 (EU Directive 2009/28/CE).

Solid biofuels as pellets, woodchips, briquettes and others can contribute significantly to this aim with a growing market in the last years (Proskurina, Heinimö et al. 2015, Eurostat Accessed: March 2015). Considering their increasing utilization, European standards for solid biofuels (EN ISO 17225-series) are continuously under development by CEN technical committee 335 in order to ensure the biomass quality. The biofuels quality control is an important tool to trace type and provenance of feedstock and its physical and chemical characteristics, which affect the efficiency of energy conversion systems and the related emissions (Toscano, Duca et al. 2014, Gehrig, Pelz et al. 2015). The chemical composition of biofuels drives the choice of conversion technology and process control in the selected energy conversion pathway (Fagan, Everard et al. 2011).

According to EN ISO 17225-series standards, a biomass is classified in a class of quality rather than in another one, according to a number of chemical-physical characteristics and quality features (i.e. origin and source) that have to meet specified limits. Some quality parameters, such as type of wood (hardwood/softwood), presence of bark or residues from wood processing industry, are difficult to be achieved by conventional analysis or microscopy techniques because time- and money-consuming. As consequence, such methods don't match the real needs of bioenergy operators that often have to take decisions in a short time. This is especially true for densified biomass such as pellets. For these kind of biofuels, traceability is particularly complex since wood is milled, loosing original structure, and different materials can participate to their blend in a concealed way, causing unpleasant effects in terms of combustion efficiency and related emissions.

A valid help comes by infrared spectroscopy, a rapid, cheap and non-destructive technique that can represent an efficient alternative to the long and expensive standard methods. Thanks to this, both qualitative (origin and source) and quantitative (chemical-physical parameters) features of the raw material might be defined simultaneously with a single and fast measurement. It is a relatively new technique, widely used due to its potential to predict a broad variety of attributes in several sectors, e.g. food, pharmaceutical and agricultural industries.

Only a limited number of studies has also investigated the application of infrared (IR) spectroscopy to the field of biomass control. Some of the most relevant are reported in this chapter and in the following ones of this Ph.D. thesis. Fagan *et al.* (Fagan, Everard et al. 2011) examined the possibility to employ near infrared (NIR) spectroscopy and chemometrics to predict moisture, ash, carbon contents and calorific value of two dedicated bioenergy crops (*Miscanthus* and short rotation coppice willow) showing that moisture and calorific value prediction models had excellent accuracy, while carbon and ash models were fair and poor, respectively. Gillespie *et al.* (Gillespie, Everard et al. 2015) also predicted the parameters of moisture, ash, carbon contents and calorific value of different pellet with hyperspectral imaging instrument. Moisture and gross calorific value prediction models had good and excellent accuracy, respectively, while the ash and carbon models were deemed good and fair, respectively. Everard *et al.* (Everard, McDonnell et al. 2012) determined the accuracy of visible and NIR spectroscopy to predict gross calorific value of dedicated bioenergy crops.

Huang *et al.* (Huang, Han et al. 2009) have looked at the use of NIR as an alternative method to predict the carbon, hydrogen, and nitrogen contents and the calorific value of straw. Liu *et al.* (Liu, Ye et al. 2010) developed a predictive model using Fourier transform near infrared (FT-NIR) spectroscopy coupled with multivariate analysis. The aim is the chemical prediction of glucan, xylan, galactan, arabinan, mannan, lignin, and ash concentration in two types of biomass feedstocks (corn stover and switchgrass).

Some studies were referred also to the real time prediction of biomass parameters during conversion processes. Lestander *et al.* (Lestander, Johnsson et al. 2009) investigated the usefulness of online NIR spectroscopy in a pelletizing process for predicting moisture content, sawdust blends and energy consumption of the pellet press. Lestander *et al.* (Lestander, Finell et al. 2012) tested NIR spectroscopy also in a newly-constructed pellet plant. The factors analysed were species composition and moisture content of the dried wood particles before pelletizing.

Infrared spectroscopy and chemometrics have been also used to get qualitative information about wood, but the number of studies is limited and only few papers can be found for the solid biofuels sector. While several authors investigated the possibility to discriminate between hardwood and softwood samples (Lestander, Johnsson et al. 2009, Chen, Ferrari et al. 2010, Russ and Fiserova 2011, Lestander, Finell et al. 2012, Yang, Lu et al. 2012), no study has been carried out on bark determination in wood samples in the biomass sector or to detect residues from wood processing industry in wood powder for pelletizing process.

Summarizing, literature research shows that IR spectroscopy was already used to investigate some aspects of biomass quality control, but the number of studies is limited in comparison to other research fields. Some aspects have been already developed, but many others still have to be taken into account. In particular, the assessment of the biomass quality should focus on qualitative information concerning the origin and source of the material and the main quantitative parameters (e.g. ash content). The overall aim of this Ph.D. research activity was to assess the biomass quality using IR spectroscopy coupled with chemometrics. The study has been especially focused on the possibilities and the limitations in applying IR spectroscopy for detailed biomass control.

1.2 Outline of the thesis

The thesis is mainly based on six articles either published or under review for publication by peer-reviewed journals and one manuscript under finalization.

The main focus is the development of an innovative analytical methodology based on infrared spectroscopy and chemometrics to characterize solid biofuels from a chemical-physical and qualitative point of view.

A brief description of the outline of this thesis follows.

First of all an overview of the state-of-art of the biomass sector is presented (Chapter 2). In particular, the current system of quality control, its problems and bottle-necks are described.

In Chapter 3, infrared spectroscopy is presented as a possible solution to the problems explained in the previous chapter. The fundamentals of medium infrared (MIR) and near infrared (NIR) spectroscopy technique are provided. Furthermore, the chapter describes how information from spectral data can be got using multivariate data analysis and chemometrics. The main multivariate analysis techniques used in this thesis will be described shortly.

Chapters 4 and 5 illustrate the experimental results of applying infrared spectroscopy in the biomass sector. The main quality key-parameters of solid biofuels, such as calorific value, ash content and moisture content, and the origin and source of the material, as defined by the standards EN ISO 17225-series, have been considered.

Finally, Chapter 6 contains a summary of the reached results and perspectives and ideas for future research.

The main manuscripts written during the three-years Ph.D. are appended at the end of this book. Paper VII is not included since it has not been submitted at present time.

2. Biomass quality

2.1 Biomass: general remarks

Biomass covers organic, non-fossil material of biological origin, which can be converted to fuel – or more precisely biofuel considering the origin of the material – for heat and electrical generation and therefore is considered a potential renewable energy source (EN ISO 17225-1:2014). As consequence it can include products, by-products and residues from agriculture, forestry and relative industries, as well as the non-fossil, biodegradable parts of industry and municipal solid waste, planted biomass and aquatic biomass (Long, Li et al. 2013). Biomass can be divided into several categories according either to their traded forms or their origin and source. The latter divides biomass into woody biomass, herbaceous biomass, fruit biomass, aquatic biomass and blends and mixtures (EN ISO 17225-1:2014). Table 2-1 gives a categorization of biomass resources according to their origin and source.

The components of lignocellulosic biomass are mainly three polymers: cellulose, hemicellulose and lignin. Cellulose and hemicellulose are polysaccharides. In detail, cellulose $(C_6H_{10}O_5)_n$ is composed of hundreds of linked glucose molecules. The cellulose chains are connected together by hydrogen and van der Waals bonds causing the cellulose to be packed into microfibrils. Microfibrils are covered by hemicellulose and lignin. The cellulose consists in crystalline and amorphous cellulose, characterized by a well-ordered structure and a randomly ordered structure, respectively (Xu, Yu et al. 2013, Bajpai 2016). Hemicellulose contains a much more complicated structure and linkages than cellulose, consisting of more than one type of monosaccharide unit. Depending on the variety of biomass, hemicellulose may contain xyloglucan, xylan, glucomannans, galactoglucomannans, etc. (Xu, Yu et al. 2013, Bajpai 2016). Lignin has a very complicated and large molecular structure, too. It is made of phenolic polymers that consist of three types of phenylpropane units: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (Xu, Yu et al. 2013, Bajpai 2016).

Table 2-1: categorization of origin-based solid biofuels. Table adapted from (EN ISO 17225-1:2014).

Categories	Materials
Woody biomass	Forest, plantation and other virgin wood
	By-products and residues from wood processing industry
	Used wood
	Blends and mixtures
Herbaceous biomass	Herbaceous biomass from agriculture and horticulture
	By-products and residues from food and herbaceous processing industry
	Blends and mixtures
Fruit biomass	Orchard and horticulture fruit
	By-products and residues from food and fruit processing industry
	Blends and mixtures
Aquatic biomass	Algae
	Water hyacinth
	Lake and sea weed
	Reeds
	Blends and mixtures
Blends and mixtures	Blends
	Mixtures

The amount and composition of these constituents in the feedstock can vary greatly due to the complex and heterogeneous nature of the biomass and this influences the choice of the optimum energy conversion technology for each type of material (McKendry 2002, Sluiter, Ruiz et al. 2010). Indeed, the inherent variability of the lignocellulosic biomass deals to the problem of verifying the properties before its use.

Among the different types of lignocellulosic feedstock, this thesis will focus mainly on pellet and woody biomass because of their widespread use as biofuels, the limited weather-related dependencies and several advantages compared to the other biomasses.

Woody biomass is present more or less everywhere, is available in many forms (dedicated and residual biomass) and can be easily stored (Spinelli, Nati et al. 2011, González, Riba et al. 2015). In particular, woodchip is really appreciated because it consists of homogeneous particles with a specific size and it guarantees benefits in terms of increased load density and handling quality (Manzone 2016). In different European countries, and in Italy as well, the number of power plants fuelled with woodchip is increasing and accordingly also the demand for wood fuels (Toscano, Duca et al. 2016).

Pellet is one of the most competitive among the different biofuels, because it presents a higher energy density, a reduced transportation and storage cost and it is durable. Different materials can participate to its blend making it adaptable to different location and feedstock (Obernberger and Thek 2010, Mola-Yudego, Selkimäki et al. 2014). The pelletizing process creates a product with regular cylindrical shape and low moisture content, making possible its use not only in heat and power plants, but also for automatic feeding in small-scale applications (Selkimäki, Mola-Yudego et al. 2010). Moreover, pellet ensures a clean burning and a reduction in ashes and emissions. For all these reasons, pellet is becoming more and more used in different countries, especially in Europe, where the market is increasing rapidly. Italy is the third largest European producer (0.77 million tonnes), but the domestic supply side doesn't cover the demand: its consumption in 2009 was 1.1 million tonnes (Sikkema, Steiner et al. 2011).

2.2 Current biomass quality control system

The identification and characterization of the chemical-physical composition of a solid fuel is the most important step during the investigation and application of such fuel. This composition is a unique code that characterizes and determines the properties, quality, application perspectives and environmental problems related to any fuels (Vassilev, Baxter et al. 2010).

In fact, it is known that there is an inherent variability in the properties of lignocellulosic material that is influenced by many factors (Dare, Gifford et al. 2001, Hames, Thomas et al. 2003) and this leads to the need to employ standards in order to check the quality of the product. Moreover, considering the increase in biomass demand, the quality could experience a decrease and need to be analysed (Toscano, Duca et al. 2016). Pellet is an example of that. As a results of the rapid growth of pellet market, different raw materials can be used to make up the product apart from woody biomass, making its traceability particularly complex. While, in the past, pellets were mainly made up by pure wood (sawdust, chips or shavings), nowadays - especially in some countries - the supply of pure wood is no more sufficient and other raw materials were used (Toscano, Riva et al. 2013), causing unpleasant effects and influencing the combustion behaviour and the related emissions. As consequence, it is very important to analyse its quality, as it has strong implications for environmental, technical and economic aspects (Toscano, Riva et al. 2013).

Biofuel quality control might then be an important tool to trace composition, provenance, chemical-physical characteristics of the raw materials, which in turn affect the efficiency in energy conversion systems and the related emissions.

CEN/TC 335 has established a number of standards (EN ISO 17225 series) relating to different solid biofuels, to ensure their quality (Figure 2-1). EN ISO 17225 under the general title "Solid biofuels – Fuel specifications and classes" consists of the following parts:

- Part 1: General requirements (EN ISO 17225-1);
- Part 2: Graded wood pellets (EN ISO 17225-2);
- Part 3: Graded wood briquettes (EN ISO 17225-3);
- Part 4: Graded wood chips (EN ISO 17225-4);
- Part 5: Graded firewood (EN ISO 17225-5);
- Part 6: Graded non-woody pellets (EN ISO 17225-6);
- Part 7: Graded non-woody briquettes (EN ISO 17225-7).

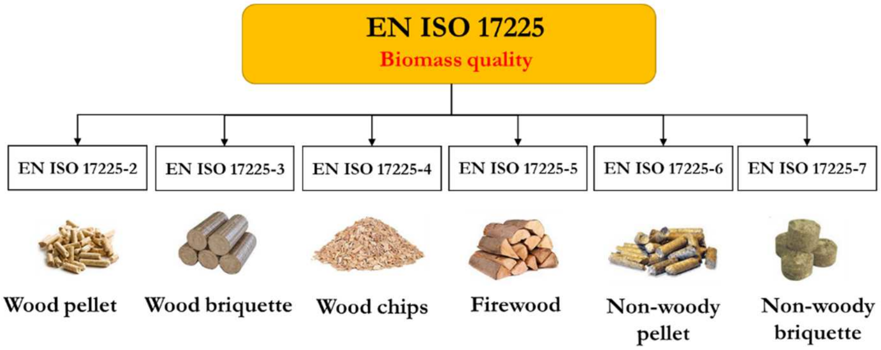

Figure 2-1: standards EN ISO 17225 series relating to different solid biofuels.

The standards divide the biomass in different quality classes – for example A1, A2 and B for woody pellet – indicating specific limits of chemical-physical parameters and defining attributes like the origin and source of the material.

Setting out ranges for several chemical-physical parameters that are relevant to assess biomass quality, the standards indicate also the procedures to determine them. These features are normally achieved by conventional laboratory analysis.

In order to ensure the accuracy of the analysis performed and of the returned results, a sampling procedure is required. Hence, a good quality control is straightly linked to a sampling process, especially when considering large bulk of biomass (e.g. plant biomass pile, biomass shipload, feedstock supply in pelletizing/power plant). Standard EN ISO 18135:2017 – Sampling of solid biofuels (EN ISO 18135:2017) – describes the procedure for performing sampling plans and taking samples representative of the whole lot of interest to be delivered to laboratories for quality control. First of all, it is necessary to collect a number of sub-samples depending on the mass of the lot and the heterogeneity of the material. Obviously, a higher number of sub-samples is needed with increasing mass of the lot and degree of heterogeneity. The sub-sample extraction must be carried out with suitable instruments, able to take one sub-sample at a time with specific volumes. The standard defines also the points where to collect the sub-samples and these are often hard to be reached by the operators, such as in big biomass pile. The sub-samples are then collected together to form a sample which is immediately closed in a sealed container in order not to alter the characteristics to be analysed.

Once arrived at the lab, the biomass sample needs to follow a specific workflow in order to define the right quality class. A summary of the crucial steps to be performed is presented in Figure 2-2.

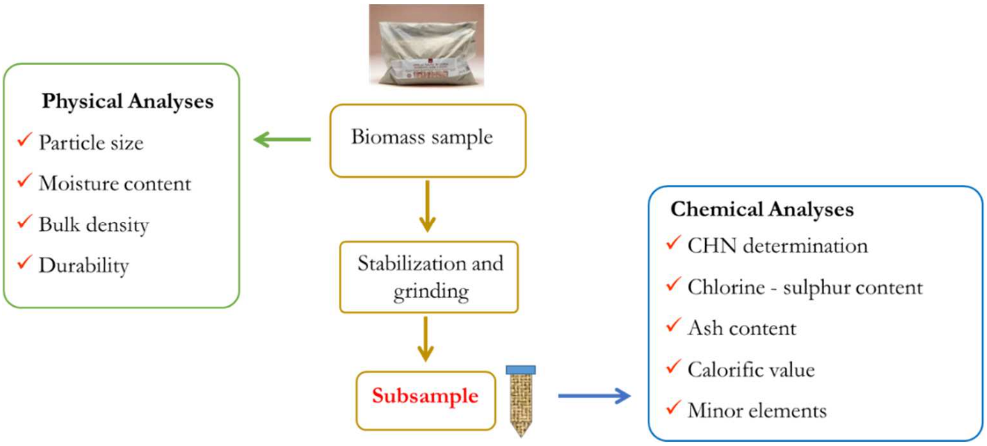

Figure 2-2: workflow of laboratory analyses for the control of biomass quality. Physical analyses are performed on the biomass sample as received at the lab and chemical analyses on the stabilized and ground material.

First of all, according to UNI EN 14780:2011 standard, part of the material is stabilized at 45 °C for 24 h then ground down to 1 mm of particle size. It is important that the representativeness of the sample, as arrived at the lab, shall not be changed during the sample reduction phase. The sample preparation phase is necessary before the samples are tested for chemical analyses, instead physical analyses are performed on the sample as arrived at the lab without any previous preparation. Some of the main physical analyses consist of particle size, moisture content, bulk density and durability. The moisture content (Mc) is the most important physical characteristic of wood fuels, it depends on different factors (Spinelli, Ivorra et al. 2011) and it influences the net calorific value.

Regarding chemical analyses, the most relevant parameters taken into consideration in the biofuel sector are ash content (Ac) and gross calorific value (GCV). The ash content is one of the most important parameter. It influences combustion efficiency and may cause cleaning and combustion problems, such as slagging in furnaces, fouling of heat exchanger surfaces (Kortelainen, Jokiniemi et al. 2015), corrosion in the combustion device (Werkelin, Skrifvars et al. 2010) and fly ash emissions (Narodoslawsky and Obernberger 1996). The calorific value is a property of a fuel, which determines the energy value of it, and it is relatively constant for wood fuels in their dry status (Erol, Haykiri-Acma et al. 2010). It is an essential parameter for the planning and control of power plants using biomass fuel (Posom, Shrestha et al. 2016). It is important for setting the price of the product and its value is influenced by the moisture content.

The analytical methodologies adopted for the determination of GCV, Ac and Mc refer to the standards EN ISO 14918:2010, EN ISO 18122:2015 and EN ISO 18134-2:2015, respectively. In particular, Mc is calculated from the loss in mass of the sample after it was dried at temperature of 105 ± 2 °C in air atmosphere until constant mass is achieved. The Ac is determined incinerating the sample to a controlled temperature of 550 ± 10 °C using a muffle furnace and calculating the mass of the residue remaining after the sample was heated. The gross calorific value is determined in a dynamic mode at 25°C using a bomb calorimeter.

Summarizing, according to EN ISO 17225, the assignment of the quality class of a biomass is based on the origin and source (qualitative information) and the chemical-physical characteristics (quantitative information) of the material. However, this current system of quality control presents some limits.

As already described, the chemical-physical parameters are determined with laboratory analysis and considering the procedures defined by the technical standards, but the process turns out to be tedious, long, expensive and requires specialized experts. As a result, these complex analyses are normally not in line with the real necessities of the sector operators involved in the quality control that need to have results as quick as possible in order to make their decision and ensure the biomass quality.

Getting information about the origin and source of the material is particular relevant, such as type of wood (hardwood/softwood) and presence of bark or by-products of wood processing industry (§ Table 2-1). This applies especially to the pellet sector, where the wood is unstructured and milled and different raw material could be used in the pelletizing process, causing problems with traceability and quality trend of the product. These attributes could be determined with laboratory analyses (e.g. microscopy techniques and chemical analysis) but, besides the same disadvantages already listed for the assessment of chemical-physical parameters, no standardized method exists for these features.

Furthermore, the structural and chemical complexity and the heterogeneous nature of the lignocellulosic material lead to the necessity of performing a correct sampling in order to have a proper quality characterization. As already mentioned, the standard EN ISO 18135:2017 – Sampling (EN ISO 18135:2017) – shows the procedure to perform a correct sampling plan and to obtain a sample representative of the whole lot of interest. The principle is based on the fact that every particle in the lot or sub-lot should have the same probability to be included in the sample. This procedure is normally quite long and hardworking and the quality parameters of interest, e.g. calorific value, moisture and ash contents, are determined using only some grams of material compared to the tons of biomass transported by ship or trucks or in input in a biomass plant. Moreover, it is important to point out that sampling procedure, which is the fundamental step to obtain valuable analytical results, requires specialized operators. In fact the employment of non-specialized staff might result in probable errors of the procedure. As a consequence the operators of the biofuel chain (power plants, producers or distributors) have two possibilities: i) to train their own personnel, ii) to assign the sampling at an external laboratory, with an increase in analytical costs. In any case, both solutions are expensive and long.

2.3 Toward a new system of biomass control

In summary, the current system of biomass quality control is based on specific ranges of chemical-physical parameters and qualitative attributes (origin and source) of the material defined by the international technical standards EN ISO 17225. However, this system presents some limits:

- there are no good standardized procedures set by the standards to determine information about the origin and source of the material with laboratory analyses;
- laboratory analyses that are used to define the chemical-physical parameters required by the technical standards are long and expensive;
- biomass exhibits an inherent variability, so it is difficult to perform a correct sampling and be representative of the total biomass.

As a consequence, a technique that is rapid, economic and simple is necessary. A good candidate could be the infrared spectroscopy which is already widely used for quantitative and qualitative purposes in different fields, such as the agricultural, food, pharmaceutical and industrial sectors. The methodology is shortly described in the following chapter (§ Chapter 3). IR spectroscopy is a suitable technique to overcome the biomass sampling and the related problems with its representativeness. In fact, it could be applied directly in the production line providing continuous measurements and giving back the quality trend of the product and the possible critical points, in real time. Beside on-line/in-line measurements, infrared technique could be used also off-line or at-line. The former includes sampling and a sample transfer to the lab that will return results after the infrared analysis, reducing in this way time and cost compared to the conventional analysis, but losing real-time knowledge about the process. The at-line measurements also require a sampling, but the sample is analysed close to the process reducing the time delay of lab analysis. To take into account that even if these techniques are still based in sampling, the problem of the representativeness is overpassed since multiple samples can be analysed thanks to the low cost and the high speed of the infrared analysis.

In general it can be concluded that IR spectroscopy technique offers different benefits. First of all, contrary to the regular analysis, it is fast, can help to achieve cost-savings and does not require skilled operators, once the method has been developed. And more importantly, it could be considered the first methodology able to provide rapid responses about biomass traceability (origin and source). All these features allow to overpass the limits of the current system of quality control.

A schematic representation of the two different approaches in solid biofuels analysis is reported in Figure 2-3: the current system of biomass quality control based on conventional analysis and the infrared system.

Paper I, **Paper** II, **Paper** III, **Paper** V and **Paper** VI aim at collecting information about the origin and source of the biomass. Instead, **Paper** IV aims at predicting two quantitative characteristics of woodchip samples: gross calorific value and ash content. In particular, **Paper** VII focuses on the prediction of ash content and investigates if scattering effects the prediction of such parameter. A preliminary study was also carried out in order to predict the quantitative characteristics of the lignocellulosic material directly in the production line using an on-line spectrophotometer (§ Chapter 5.3). More precisely the parameters considered are moisture content on woodchip, ash content and gross calorific value on pellet samples.

Going into detail, the purpose of **Paper** I is the discrimination between softwood and hardwood samples. Once achieved this objective, **Paper** II aims at predicting hardwood and softwood contents in blend samples. **Paper** III and **Paper** V and VI investigate other issues related to the origin and source of the biomass: the discrimination between bark and wood samples and between treated and virgin wood, respectively.

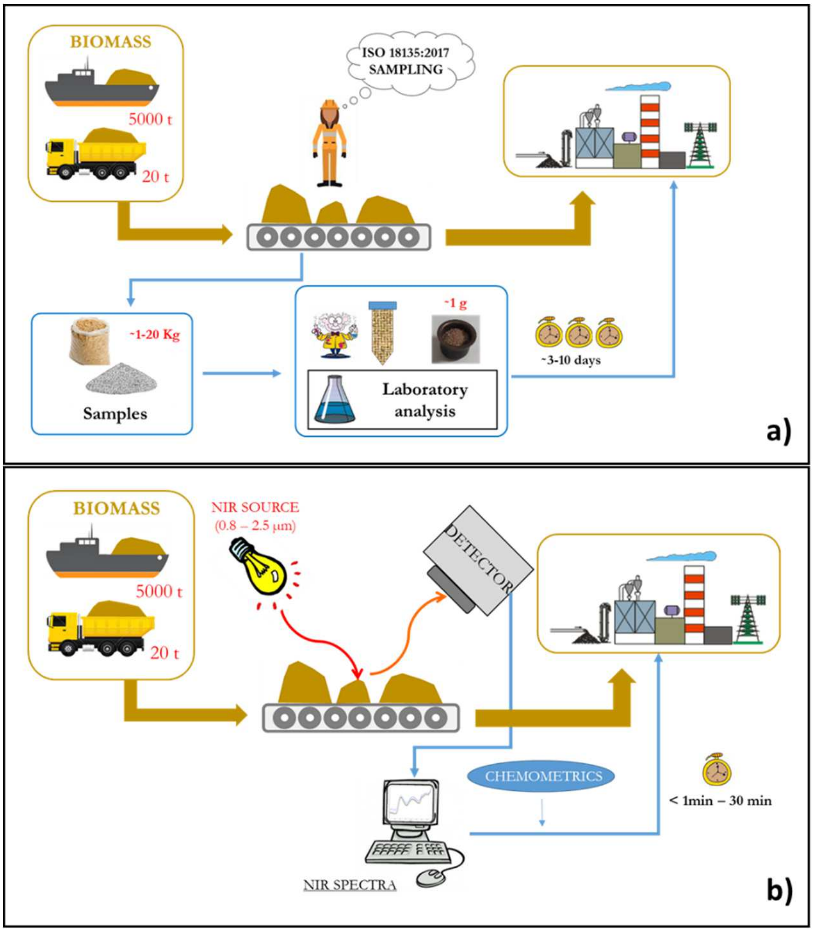

Figure 2-3: different approaches of solid biofuels control: the current system of quality control based on conventional analysis (a) and the infrared system (b).

3. Infrared spectroscopy and chemometrics

3.1 Fundamentals of infrared spectroscopy

Spectroscopy is an analytical technique that studies the interaction between matter and different frequency components of the electromagnetic spectrum. Infrared spectroscopy is the spectroscopy that deals with the infrared region of the electromagnetic spectrum. In detail, the infrared region is typically divided into the NIR – near infrared (12,500-4,000 cm^{-1}), the MIR – medium infrared (4,000-400 cm^{-1}) and the FAR – far infrared region (400-100 cm^{-1}) (Figure 3-1) (Pavia, Lampman et al. 2001). This thesis will focus on the NIR and MIR regions because of their utility and application in the sector.

In the NIR region absorption bands correspond mainly to overtones and combinations of the fundamental vibrations of molecules observed in the MIR region (Blanco and Villarroya 2002, Siesler 2007).

When infrared radiation strikes a sample, the molecules can absorb energy from the electromagnetic radiation and convert it into energy of molecular vibration. The vibrational energy refers to an oscillation of atoms through their molecular bonding. Chemical bonds of the molecule vibrate at different energy levels depending on the molecular structure, chemical composition or physical properties of the surface measured. There are two types of vibrational mode, consisting in an elongation of the chemical bonds (stretching) and/or in a deformation of the angle of the chemical bonds (bending).

The absorbance of light by the molecular vibrations is linked with a change in the polarity of the bonds (dipole moment). In fact, the transfer of energy occurs only if the electric dipole of the vibrating bond changes at the same frequency of the incoming radiation (Siesler 2007).

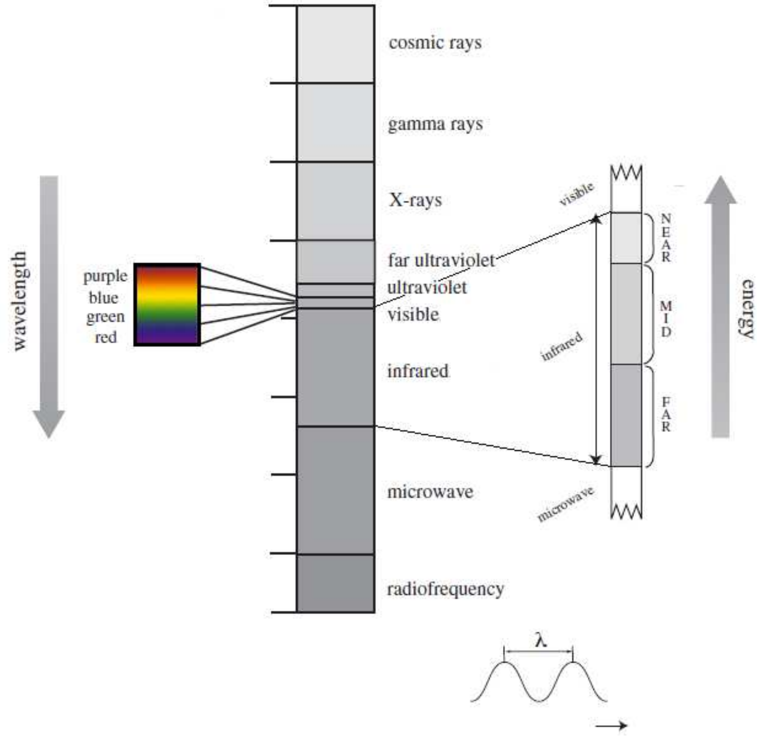

Figure 3-1: the electromagnetic spectrum. Illustration adapted from (Department of Chemistry and Biochemistry 2002).

The key-point of the spectroscopic measurements lies in the association between the frequency of the radiant energy and the frequencies of molecular motions. Since only certain light frequencies or wavelengths can be absorbed by a specific molecular vibration, spectroscopy can be used to get qualitative information about the material analysed. Quantitative information could be also acquired using infrared spectroscopy based on the amount of light absorbed at each wavelength.

To better understand the vibration model, a chemical bond could be viewed as a spring with a given force constant (k) connecting two vibrating masses (m). According to Hook's law, the wavenumber of the fundamental absorption band ($\tilde{\vartheta}$), i.e. the frequency of vibration, is described as (Pasquini 2003, Siesler 2007, Jha 2016):

$$\tilde{\vartheta} = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}$$

where k is the force constant of the chemical bond and μ is the reduced mass of the atoms involved in the bond, given by the following equation (Pasquini 2003, Siesler 2007, Jha 2016):

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

in which m_1 and m_2 are the masses of the two bonded atoms.

In the harmonic oscillator, energy is continuous and the system undergoes cyclic conversion from potential energy to kinetic energy. In reality, the molecules do not follow the rules of the classical mechanics, instead those of quantum mechanics. That means that the transition between two energy states occurs with the absorption or emission of a 'quantum' of energy, i.e. a discrete packet of energy. Hence, the absorption or emission of energy by the system causes a transition between two energy levels – $E1$ and $E2$ – and the energy difference (ΔE) between two vibrational states is equal to $\Delta E = h * \vartheta$ where h is the Planck's constant and ϑ is the frequency (Pasquini 2003, Siesler 2007, Jha 2016). Since vibrational energies are quantized, the vibration frequencies are calculated as following (Pasquini 2003, Siesler 2007, Jha 2016):

$$\Delta E = h\vartheta\left(n + \frac{1}{2}\right)$$

in which ϑ is the frequency, n is the vibrational level and h is the Planck's constant (Figure 3-2) (Department of Chemistry and Biochemistry 2002). To summarize, the molecular vibration can be described by a harmonic oscillator, in which the energy levels (n) are equidistant and transitions are only allowed between neighbouring energy levels with $\Delta n = \pm 1$.

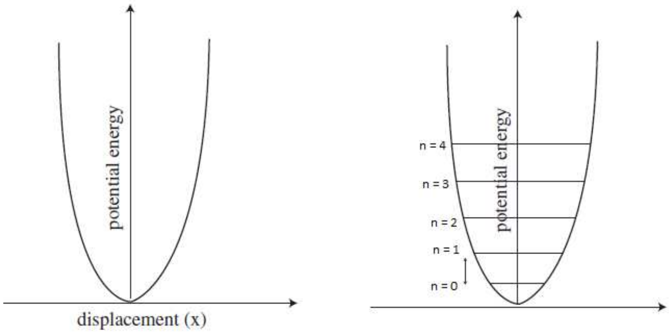

Figure 3-2: energy curve for a vibrating dipole molecule, classic mechanics (left) and energy constrained to quantum mechanics (right). Illustration adapted from (Department of Chemistry and Biochemistry 2002).

The quantum-mechanical treatment of a harmonic oscillator explains the infrared absorption bands caused by the fundamental vibration, but do not explain the overtone bands in the NIR region. The bonded atoms are compared to two masses linked by a spring, but in reality it is also necessary to consider the dissociation energy of the bonds. The distortion of the equilibrium of internuclear distance, caused by the strongly extension or compression beyond a certain point of the vibrating bond, leads to bond dissociation. Moreover, when two atoms approach, the repulsion between the two nuclei occurs and the potential energy increases more quickly than that estimated by the harmonic oscillator (Jha 2016). The harmonic oscillator becomes an anharmonic oscillator (Figure 3-3) (Department of Chemistry and Biochemistry 2002).

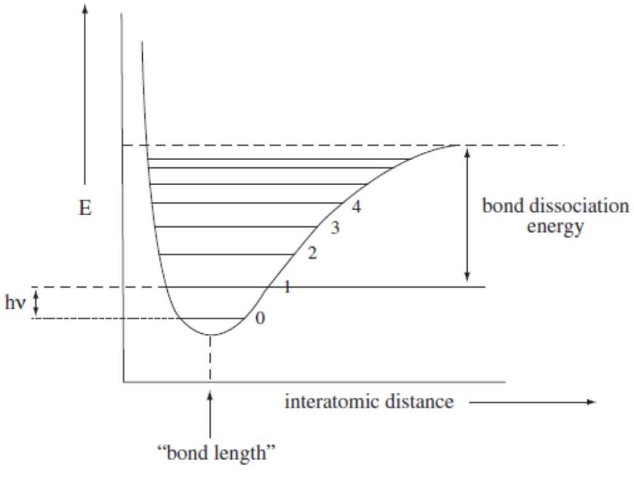

Figure 3-3: energy curve of an anharmonic oscillator. Illustration adapted from (Department of Chemistry and Biochemistry 2002).

In the case of NIR spectroscopy the position of an overtone is multiple of that of the fundamental vibration, as consequence $n > 1$. For this reason, the position of the overtones can be found as described in the formula:

$$\vartheta_n = n\tilde{\vartheta}(1 - n\aleph)$$

in which ϑ_n is the wavenumber of the overtone corresponding to the vibrational quantum number (n), $\tilde{\vartheta}$ is the frequency of vibration and $\aleph$ is the anharmonicity constant (Eskildsen 2016).

3.2 Spectrophotometers

Spectrophotometers are the instruments used to collect spectra, normally visualized in a graph of infrared light absorbance on the vertical axis and wavelength on the horizontal axis. The components of the instrumentation include a radiation source, a dispersing element (or wavelength selector), a detector and a data processing unit. The dispersing element is used to define the wavelength selection, the way in which light is spectrally modulated or selected. Different technologies are available, making it a crucial component in the choice of the infrared system (Blanco and Villarroya 2002, Xu, Yu et al. 2013). Mainly all the instruments are grouped into dispersive and interferometric.

In **Paper** I and in **Paper** II, Nicolet iS 10 (Thermo Fisher Scientific Inc, USA) equipped with an attenuated total reflectance accessory (mod. Smart iTR, Thermo) was used for obtaining MIR measurements. In **Paper** III, in **Paper** IV and in **Paper** VII, Antaris II (Thermo Fisher Scientific Inc, USA) was used for obtaining NIR measurements. In **Paper** V and **Paper** VI, Nicolet iS 10 (Thermo Fisher Scientific Inc, USA) equipped with a diffuse reflectance accessory (mod. Smart Near-IR upDRIFT, Thermo) was used for obtaining NIR measurements.

All the spectrophotometers are based on the use of the interferometer as wavelength selector system and on Fourier transform (FT) principle. The Michelson interferometer consists of two mutually perpendicular plane mirrors, one of which moves along its axis at a common velocity, and a beamsplitter (Figure 3-4 a) (McCarthy and Kemeny 2007). The energy of the source is collimated and directed into the beamsplitter, generating two equal beams. One beam is reflected towards the fixed mirror and the other beam toward the moving mirror. Light is reflected from the two mirrors back to the beamsplitter and part of the light passes into the sample and the detector records the signal. However, there is a difference in the optical path length of the two beams (retardation) because of the continuous change caused by the motion of the moving mirror. The result is an interferogram where the absorbance values are recorded in the time-domain since the signal is recorded by the detector for various values of the retardation. The application of Fourier transform principle allows the transformation of the interferogram into the familiar frequency-domain spectrum (normal spectrum) (Pavia, Lampman et al. 2001).

The interferometer is an efficient method to enable measurements of all infrared frequencies simultaneously, rather than individually. They include the best characteristics in terms of wavelength precision and accuracy, high signal-to-noise ratio and scan speed (Pavia, Lampman et al. 2001, Pasquini 2003), essential also for in-line applications.

A part from interferometric, the other common technology used to produce IR spectra is the dispersive optics-based instrument. These instruments are based on diffraction gratings and thanks to recent evolution in sensors production technology, they allow a higher scan speed than the past. That is because it is now possible to construct linear arrays of PbS and InGaAs sensors containing up to 256 independent elements. The procedure is simple: the incoming radiation is dispersed by a diffraction grating that separates the wavelengths of light and directs each wavelength individually through an array of detector elements (Figure 3-4 b). The pros of this technology is the high scan speed, in fact thanks to the detector array it is possible to scan an entire spectrum in a few milliseconds, and the absence of moving parts, allowing the installation for in-line applications (Pasquini 2003, Jha 2016).

In this thesis an on-line NIR spectrophotometer loaned by Büchi (NIR-Online; Büchi Labortechnik AG, Flawil, Switzerland), based on diode-array technology, was used for collecting NIR spectra in order to investigate the possibility to predict the biomass quality parameters directly in the production line (§ Chapter 5.3).

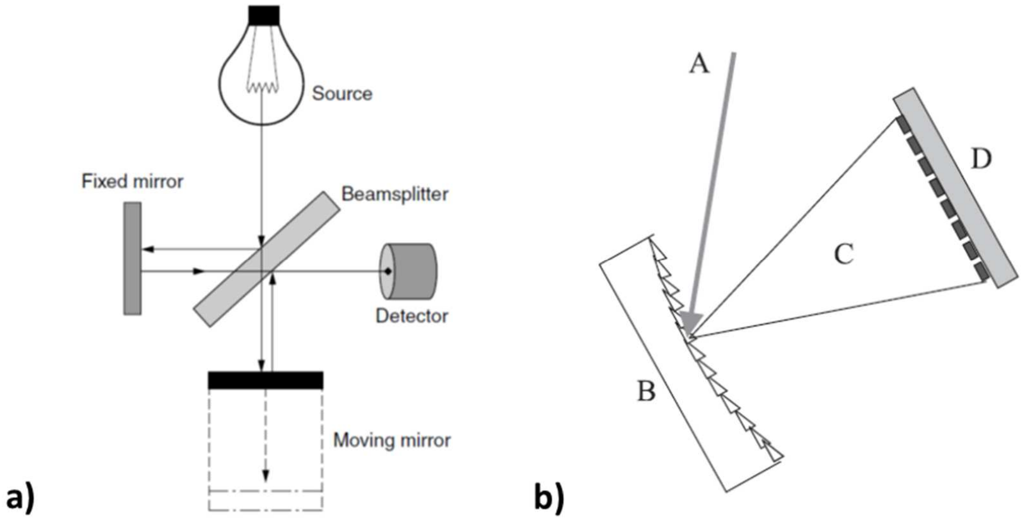

Figure 3-4: a) Michelson interferometer (McCarthy and Kemeny 2007); b) dispersive optic-based spectrophotometer. A, incoming radiation; B, concave grating; C, dispersed radiation, and D array of detectors (Pasquini 2003).

3.3 Chemometrics

Spectroscopy measurements are very easy and rapid to perform and allow you to get thousands of data points in few seconds, but the spectra obtained are quite difficult to interpret, even if they are very rich in inherent information. Especially the overtone and combination bands of the NIR region are weak, very broad and overlap heavily, leading to spectra really complex to interpret and with spectral differences between samples very minimal and difficult to find by naked eye (Blanco and Villarroya 2002, Lavine 2003, Hindle 2007).

The interpretation of the spectral information is possible using chemometric techniques. Chemometrics is the science of extracting information from chemical systems by using several mathematical and statistical methods based on multivariate data analysis (MVA).

First of all, one must consider the fact that the absorption signal is not only affected by the chemical properties of the sample, but also by the physical properties. In fact, a typical physical phenomenon during spectra acquisition consists in the deviation of light from a straight trajectory by one or more paths with no absorption. This is called scattering and it is viewed in infrared spectroscopy as offset and slope differences. When physical properties are not of interest, it is necessary to remove this baseline shift using pre-processing methods on the spectral data prior to data analysis. A description of the most important pre-processing methods is carried out by Rinnan *et al.* (Rinnan, Berg et al. 2009). Pre-processing techniques can be divided in two categories: scatter correction methods and spectral derivatives.

The first group includes Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV). The MSC correction is achieved by regressing a measured spectrum against a reference spectrum and then correcting the measured spectrum using the correction coefficients (slope and intercept) of this reference spectrum. In most applications, the average spectrum of the calibration set is used as the reference spectrum. Practically, the MSC technique consists in:

$$\mathrm{x_{corr}} = \frac{\mathrm{x_{org}} - b}{a}$$

in which $\mathrm{x_{corr}}$ is the spectrum corrected with MSC, $\mathrm{x_{org}}$ is the measured spectrum, b is an additive correction factor (intercept) and a is a multiplicative correction factor (slope). a and b are found by the regression of $\mathrm{x_{org}}$ onto $\mathrm{x_{ref}}$, where the latter is the reference spectrum.

SNV is a weighted normalization, i.e. not all points contribute to the normalization equally. In SNV correction, the average value of the spectrum is subtracted from the absorbance for every data point and the result is divided by the standard deviation. Hence, the entire sample is normalized by the standard deviation. Different vector-norms could be used for the scatter-correction computation, but surely standard deviation is the most applied.

SNV uses the same signal-correction concept as MSC, but a reference signal is not required, so each observation is computed on its own, isolated from the rest of the dataset. In any case, MSC and SNV are the same for most practical applications and the results are similar (Lavine 2003, Rinnan, Berg et al. 2009).

The basic idea behind MSC has been extended into more elaborate augmentations, one of the most common is the Extended Multiplicative Signal Correction (EMSC). This pre-processing method is used for eliminating uncontrollable path length or scattering effects and is particular useful in minimizing wavelength-dependent light scattering variation (Martens and Stark 1991).

This extension includes both second-order polynomial fitting to the reference spectrum (MSC normally uses the first-order polynomial fitting, i.e. the linear regression), fitting of a baseline on the wavelength axis, and the inclusion of *a priori* knowledge from the spectra of interest or spectral interferents (Rinnan, Berg et al. 2009). This last extension is not always possible since, in many practical cases, reference spectra for wanted or unwanted constituents are not available. Wavelength-axis dependency considers most often a second-order polynomial fitting on the wavelength axis to the spectra.

As a consequence, the basic equation of MSC can be expanded to include any other appropriate corrections, not only the intercept and slope already considered by MSC, but also the correction factors according to i^{th} order of the reference and the i^{th} order wavelength-axis dependency. It is thus clear that EMSC is nothing else that an expansion of the original form of MSC.

EMSC is particularly used because of its flexibility in performing a more selective correction for various types of scattering and other unwanted variations. Moreover, it is a reliable pre-treatment in the field of IR spectroscopy for the correction of additive baseline effects, multiplicative scaling effects, and interference effects (Afseth and Kohler 2012).

The spectral differentiation has the capability to remove both additive and slope effects in the spectra. The first derivative removes only the baseline; the second derivative removes both baseline and linear trend. During the differentiation, a smoothing of the spectra is performed in order to decrease the detrimental effect on the signal-to-noise ratio that conventional finite-difference derivatives would have (Rinnan, Berg et al. 2009). In this thesis, Savitzky-Golay polynomial derivative filter (Savitzky and Golay 1964) was used.

After performing the right pre-treatments, MVA methods are applied in order to extract the desired chemical information from absorbance, transmittance, or reflectance data. There are several MVA methods which have been extensively described in the literature and have been classified according to their purpose and the algorithms or computational procedures that they use.

The main categories of analysis techniques are:

- *exploratory analysis*: Principal Component Analysis (PCA) is the main widespread unsupervised approach. It simply “looks” at the data to explore the overall statistical variance with the goal of clustering the samples, according to the spectral differences, and detecting outliers;

- *classification and discrimination*: allow to discriminate between groups and classify samples in different class memberships. It can be applied when some prior information about the samples is available. Partial Least Square Discriminant Analysis (PLS-DA) and Soft Independent Modelling of Class Analogy (SIMCA) are two of the most used techniques in different fields;
- *regression*: allows the prediction of the characteristics and properties of unknown samples comparing blocks of data. The most used technique is Partial Least Square regression (PLS).

A brief description of the chemometric techniques included in this Ph.D. thesis is provided in the following sections.

3.3.1 Principal Component Analysis (PCA)

Spectral data are collected in a matrix (*X*) with size *N* x *K*, where *N* are rows (observations) and *K* are columns (spectral variables). In spectroscopy, it is normally to work with huge datasets, considering a number from several hundreds to thousands of variables. PCA, invented by Karl Pearson in 1901 (Pearson 1901) and developed and named by Harold Hotelling in 1933 (Hotelling 1933), is an exploratory analysis that allows the reduction of the dimensionality of the dataset (*X*) by transforming the original variables (*K*) into a set of new uncorrelated variables called principal components (PCs). Each of these new variables retains a part of the original variance (i.e. information) and they are ordered on the base of the amount of explained variance.

Basically, it uses orthogonal transformations to search for the direction of maximum variability and it projects the variable with the greatest variance, i.e. the first principal component (PC1), on the first axis. The variable with the next highest degree of variance in the dataset is the second principal component (PC2) and is set orthogonal to PC1. The third PC (PC3) is the best rank after the first two principal components and has the constrain to be orthogonal to the first two PCs. The procedure continues till all the systematic variation in the *X* matrix is explained. Normally, the relevant information for the system is gathered in a reduced number of variables making the data easier to explore and visualize.

The outcomes of the PCA decomposition are the scores matrix (*T*), the loading matrix (*P*) and the residuals matrix (*E*). The decomposition by PCA is described by the following equation:

$$X = TP^T + E$$

where *T* and *P* represent the structure part of spectral information and *E* the noise part.

The intrinsic information contained in the *X*-matrix can visually summarized by plotting the scores matrix. The resulting graph (scores plot) allows to visualize one principal component versus another one (e.g. PC1 versus PC2 or PC1 versus PC3). This is a very informative plot used for looking at similarities or dissimilarities among the samples and for the detection of clustering or outlier samples. Samples with similar characteristics have similar scores values and are close to each other, while samples that lie in the opposite part of the plot have very different characteristics and, accordingly, different scores values.

The loading matrix is represented by using a two dimensional-plot in which each loading vector (or at least the most relevant) is plotted against the wavelength number. This plot is used for interpretation purposes since it describes the way in which the original variables are linearly combined with the new variables (PCs). The loading plot explains why samples are grouped in a certain way in the scores plot and which variables are responsible for such separation.

A schematic representation of PCA structure is reported in Figure 3-5.

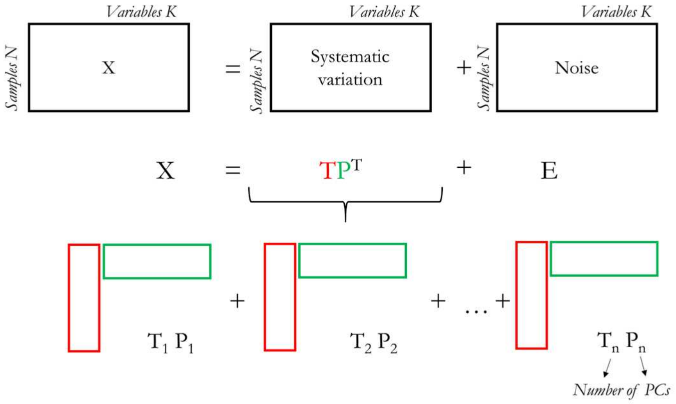

Figure 3-5: schematic representation of PCA model. T and P represent the scores and loading matrices, E is the residual matrix. X is the spectral matrix with the size N x K (samples x spectral variables). n is the number of Principal Components chosen to build the model. Adapted from (Bro and Smilde 2014).

3.3.2 Partial Least Square regression (PLS)

Partial Least Square regression (PLS) is a multivariate technique applied in spectroscopy for quantitative analysis (Wold, Sjöström et al. 2001). It finds the relationship between a *Y*-value – the parameter to be quantified – and the spectral data matrix (*X*), maximizing the covariation between *X* and *Y* scores. In short, PLS finds components in such a way that the score values in *X* and in *Y* have maximum covariance. The model is then built only using the *Y*-related variance in *X*.

As for PCA, PLS finds a new smaller set of variables, called latent variables (LVs), that are linear combinations of the spectral data and relevant for the determination of the parameters of interest.

PLS decomposition could be written with the following equations:

$$X = TP^T + E$$

$$Y = UQ^T + F$$

where *X* is the spectral data matrix (matrix of predictors) and *Y* is the parameter to be quantified (vector of response), *T* and *U* are the *X* scores and the *Y* scores matrices, respectively, *P* and *Q* are the *X* and *Y* loading matrices, respectively, and *E* and *F* represent the residuals.

After building a PLS model (calibration phase), it must be validated in order to evaluate its prediction performance (validation phase). The validation computations can be made on a separate set of data (test set) or on the calibration data themselves simulating a prediction testing. The former procedure is called external validation and the latter cross-validation. It is based on deleting a sample or groups of samples (*g*), building a model with the remaining *N* – *g* samples, estimating the coefficients of the equation and predicting the deleted sample/s. In order to estimate the uncertainty of the model RMSE (Root Mean Square Error) is calculated. It is the square root of the average sum of squares of the differences between measured values and the model estimated values for the calibration samples. The validation phase allows the computation of RMSECV (Root Mean Square Error of Cross-Validation) if using cross-validation and RMSEP (Root Mean Square Error of Prediction) if using external validation. It is the most important error measure to estimate the performance of the regression model and is nothing else than the average error that it can be get during prediction, i.e. the average error in the *Y*-vector in absolute value, so in the original scale of the *Y*-value.

Apart from quantitative analysis (*Y*-value contains quantitative information), PLS might be used also for classification purposes (*Y*-value contains qualitative information and is categorical). This multivariate method is called PLS Discriminant Analysis (PLS-DA) (Wold, Sjöström et al. 2001, Barker and Rayens 2003). It is one of the most well-known supervised statistical technique employed in chemometrics. This methodology is used for building classifiers and for predicting the belonging class of unknown samples, i.e. to predict a categorical variable. It can be applied when some *a priori* information about the sample is available and it assesses the relationship between a *Y*-value and the spectral data. As the *Y*-vector is a dummy vector, indicating if the sample belongs to a class or not (i.e. 1 for class A, and 0 for class B), the *Y*-predicted will have values in the range from 0 to 1, and a sample is included in a classification class rather than in another one on the basis of a defined classification rule.

The Receiver Operator Characteristic (ROC) curve is one of the most widely used statistical tool that assesses the ability of a binary classification model, defying a threshold limit on a *Y*-variable and classifying everything above that limit as admitted in a class and everything below as not admitted. Starting from a confusion matrix (Table 3-1), ROC curve is nothing else than the plot of True-Positive Rate (TPR) versus False-Positive Rate (FPR) for every possible classification threshold. By looking at the ROC curve, the optimal threshold value (and optimal classification model) can be defined (Ballabio and Todeschini 2009).

Table 3-1: confusion matrix for binary classification.

		Assigned class	
		P	N
True class	P	TP	FN
	N	FP	TN

The accuracy of the classification model is determined with sensitivity and specificity values. Sensitivity or True-Positive Rate (TPR) is a statistical measure of how well the classification model is able to recognize a sample belonging to a given class. Specificity or True-Negative Rate (TNR) measures the ability of the model to reject all the samples not belonging to that given class. The misclassification rate is also taken into account for evaluating the model performance since it is the ratio of false positives and false negatives to the total number of samples.

3.3.3 Soft Independent Modelling of Class Analogy (SIMCA)

Soft Independent Modelling of Class Analogy (SIMCA) is a supervised method of data classification, introduced by Wold and Sjöström in 1977 (Wold and SjÖStrÖM 1977). SIMCA performs a PCA for each class and builds a multidimensional space around them using the principal components calculated. Then the classification of unknown samples is made evaluating the distance between sample and models; if lower than the critical distance calculated for a specific class at suitable significance levels, then the sample belongs to that class. As class modelling technique SIMCA can also associate samples to more than one class or with no class.

3.3.4 Variable selection methods

It is well known that variables selection methods could enhance the model performance in comparison to that of the full-spectrum and give better predictions (Andersen and Bro 2010). Moreover, removing many irrelevant, noisy or unreliable variables the model complexity decreases, resulting in a model that is easier to understand. This not always leads to get better predictions, but it helps in understanding which variables contribute positively to the model (Andersen and Bro 2010). Variables selection is really relevant also for industrial on-line or at-line purposes. In this kind of applications, time and money are important key-points and often using a high-resolution instrument scanning a whole spectrum is too costly and takes too much time considering the industrial needs (Andersen and Bro 2010).

There are different variables selection methods, but the procedure is always the same. It consists in developing the model, selecting the optimal number of LVs and in re-calculating the model on the wavelengths selected on the basis of the variables selection method.

In **Paper** IV, different variables selection methods were applied in order to improve the PLS models for the prediction of ash content and gross calorific value and in **Paper** VI to enhance the classification performance between treated and virgin wood samples. In particular the methods used are: regression vector

(B), Variables Important in Projections (VIP) (Wold, Sjöström et al. 2001, Chong and Jun 2005), Selectivity Ratio (SR) (Rajalahti, Arneberg et al. 2009), interval Partial Least-Squares (iPLS) (Norgaard, Saudland et al. 2000) and recursive weighted Partial Least Squares (rPLS) (Rinnan, Andersson et al. 2014). In these studies, the best results were obtained using rPLS method.

rPLS is related to Jack-knifing, but instead of eliminating variables after each iteration, the method iteratively uses the regression coefficients to magnify important variables and thus relatively down-weight less important ones. The key-point of this procedure is that the regression vector reflects the importance of the variables. Basically, the fundamental is that the regression coefficients are recursively used as weights on the original data matrix. The procedure consists in a process of repeated PLS models and the calculated regression coefficients (*b* vector) are used as cumulative weights on *X*:

$$X_R = X_{R-1} \times diag(b)$$

where X_{R-1} is the previous updated weighted *X*, *b* is the regression coefficient from the last model (on X_{R-1}) and X_R is the "new *X*" to be used in the subsequent PLS model.

Reweighting each time the *X* matrix leads to gradually changing the *X* space, but the target vector *Y* remains unchanged. After the first weighting, some variables have been amplified and others are attenuated. In the second model based on weighted data, the amplified variables will contribute more, and the attenuated will contribute less and so on. At the end, the model will converge to a limited number of variables, but it will exhibit optimal prediction performance before, normally including co-linear neighbour variables.

A schematic representation of the rPLS procedure is reported in Figure 3-6. In rPLS, regression coefficients are represented as a diagonal matrix only for mathematical purposes.

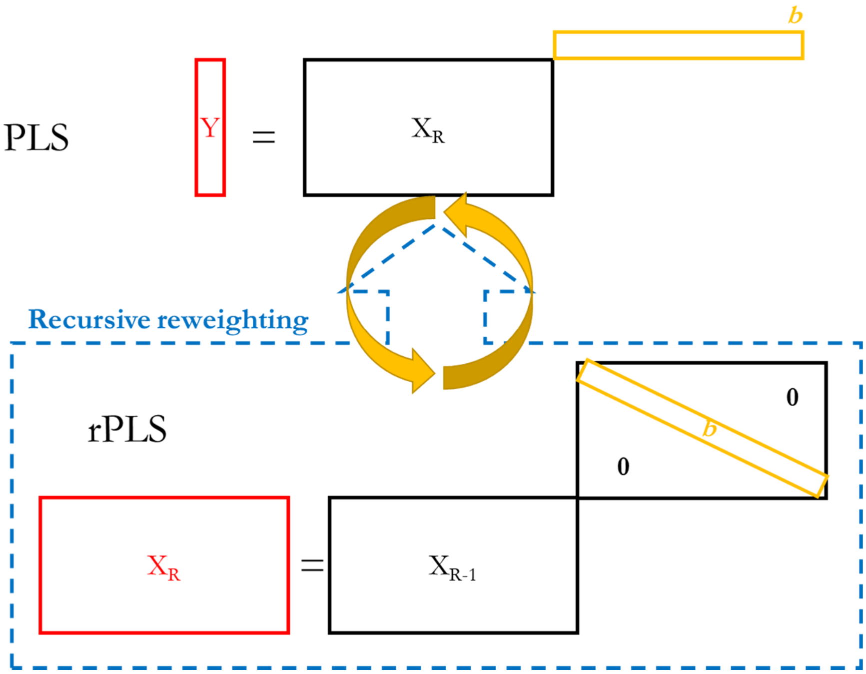

Figure 3-6: schematization of rPLS procedure. X_{R-1} *is the previous updated weighted X, b is the regression coefficient from the last model and* X_R *is the "new X" to be used in the subsequent PLS model.*

The graphical output of rPLS method is reported in Figure 3-7. The red spectrum is the average NIR spectrum, the red dotted line shows the optimal rPLS model based on the RMSECV values reported all the way to the right in the figure (this means that the threshold is not chosen but calculated on the basis of the regression/prediction error). The corresponding number of iterations needed can be read from the axis on the left ("Iteration number"). The weights are log10 transformed to enhance the view and the colour coding is reported on the left bar. The dark blue (value = 1) means that the variable has a large weight and thus important, instead the light grey (value = 1e-35) means that it is not.

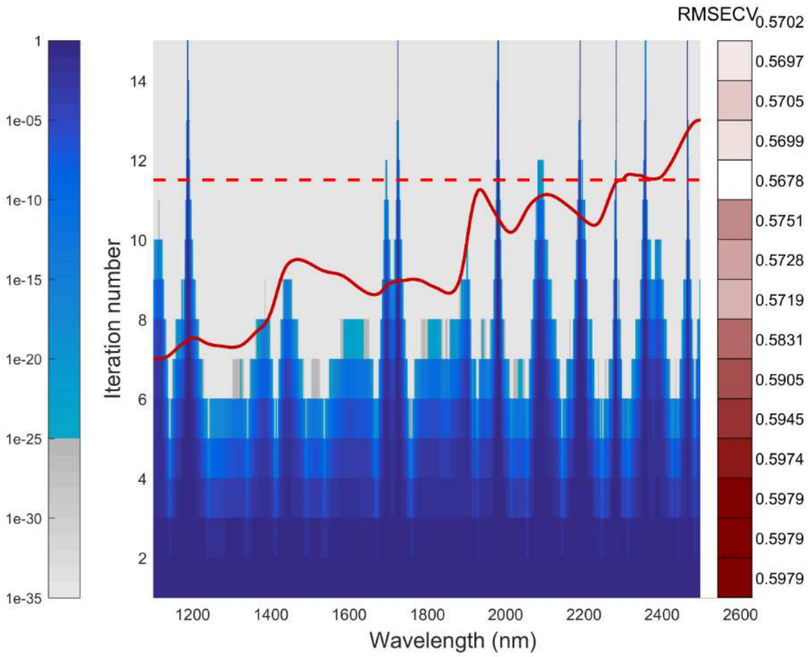

Figure 3-7: graphical output of rPLS method (Rinnan, Andersson et al. 2014).

4. Infrared spectroscopy for qualitative analysis of biomass

Getting qualitative information about the biomass is particular relevant for a number of reasons, and not only because it is required by the technical standards. Just a few examples, it allows to trace composition and provenance of the raw material, to ensure the absence of unpleasant materials (e.g. bark), to verify the compliance of producers' declarations and to assure sustainability issues. This is important especially for pellet, because the material is unstructured and milled, making impossible the determination of its origin and source by naked-eye. No expert wood technologist could be able to identify the presence of bark or other unpleasant materials mixed in the wood powder.

The great potential of infrared spectroscopy has been widely recognized both for qualitative and quantitative analysis of different organic materials, not only in the biofuel sector (Xu, Yu et al. 2013, Chadwick, McDonnell et al. 2014), but especially in other fields such as agricultural, food, industrial, textile, pharmaceutical, petrochemical, etc. (Tsuchikawa and Kobori 2015). The technique was already used to assess the qualitative parameters of the biomass. According to Tsuchikawa *et al.* (Tsuchikawa and Kobori 2015), different authors studied the utility of NIR spectra for the classification of the wood materials.

In this chapter, the main aim is to examine the possibility to use mid and near infrared spectroscopy as a tool to improve the biomass quality control. The following studies will investigate the potential to use such technique to verify the compliance of producers' declarations about wood origin and source, as defined and requested by solid biofuel quality assurance standards. In detail, it was used a) to discriminate between hardwood and softwood samples; b) to predict hardwood/softwood content in blends samples; c) to discriminate between bark and wood samples and d) to discriminate between treated and untreated wood.

4.1 Discrimination between hardwood and softwood samples *(Paper I)*

This section is mainly based on the paper "Soft Independent Modelling of Class Analogy applied to infrared spectroscopy for rapid discrimination between hardwood and softwood" accepted for publication in the peer-reviewed journal *Energy*. Part of the section is also based on the results presented with the poster "Spettroscopia nel vicino infrarosso e analisi multivariata dei dati come supporto per la determinazione della qualità del pellet" ("Near infrared spectroscopy and multivariate data analysis as a support for pellet quality determination") during the conference "7° Simposio italiano di Spettroscopia NIR" ("7° Italian Symposium of NIR Spectroscopy", Milan 13th-14th October 2016).

The main aim of the study is the discrimination between softwoods, hardwoods and the related blends by means of FT-IR and SIMCA computation. According to Clause 6 of EN ISO 17225-1 (EN ISO 17225-1:2014), solid biofuels classification is based on the origin (woody, herbaceous, fruit, aquatic, blends and mixtures) and source (virgin biomass, by-product, industry residues) of the material. As consequence, hardwood and softwood material employed for biofuel production should be declared by the producers and, in order to check the presence of non-conformity products, a rapid and cheap tool to get information about its origin and source is necessary, especially in unstructured, milled or densified wood.

For this study, a total of 110 samples have been collected. In detail, the dataset consists of 75 hardwoods of nine different species and 35 softwoods of three species. The selection of the wood species was based considering the most common species in the energy pellet sector and choosing only whole pieces of wood, like beams or boards, and debarked tree log disks wood slices. The samples collection phase was particularly important because the successive multivariate elaboration was based on the information acquired in this stage. For this reason, only samples with known origin and non-chemically treated have been selected. All the collected samples were ground, sieved under 0.25 mm particle size and analysed in medium infrared using a FT-IR spectrophotometer. A schematic representation of the workflow is reported in Figure 4-1.

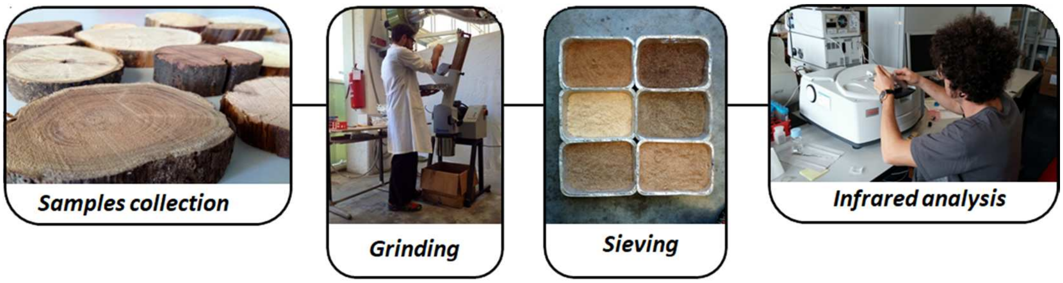

Figure 4-1: workflow of samples preparation: samples collection, grinding, sieving and infrared analysis.

The total dataset of samples was split into training set (82 samples) and test set (28 samples). The test set was chosen randomly in order to have more realistic previsions.

Besides samples, 4 blends were also analysed in the same way as samples to evaluate the discrimination performance of the developed models. The blends were prepared at 50% hardwood and 50% softwood (w/w %) selecting only the most common species of the European pellet market, i.e. beech and sessile oak for hardwood and fir and pine for softwood. In order to increase the representativeness, average samples of each common species were prepared by mixing 10 samples of the same species and using them for the blends preparation.

First of all an exploratory study (PCA) was carried out on the training set. PCA model was computed considering the *fingerprint region* from 1850 cm^{-1} up to 650 cm^{-1} and pre-treating the data with SNV and 1st derivative (second polynomial order, 13 smoothing points with Savitzky-Golay (Savitzky and Golay 1964)). The three replicates of each sample were averaged since no replicate outlier was found. An error ellipse was calculated at specific levels of significance (95%, 99% and 99.9% confidence level) both for hardwood and softwood groups. Finally, to test the ability of the PCA model in discriminating between hardwood and softwood, blends and test set samples were projected on PCs space, using the loadings calculated in PCA.

As it is shown in Figure 4-2, the separation between hardwood and softwood groups is mainly linked to PC1: softwoods have negative PC1 scores whereas hardwoods have positive PC1 scores. The higher scattering of hardwood samples in the plot is linked to the higher variability within this group. The score plot also displays that only one training set sample of hardwood and one test set sample of softwood come out from the 99.9% ellipsoid and that the projections of the 4 blends are well positioned in the middle between the two groups. All these factors suggest that the computation of SIMCA model is possible.

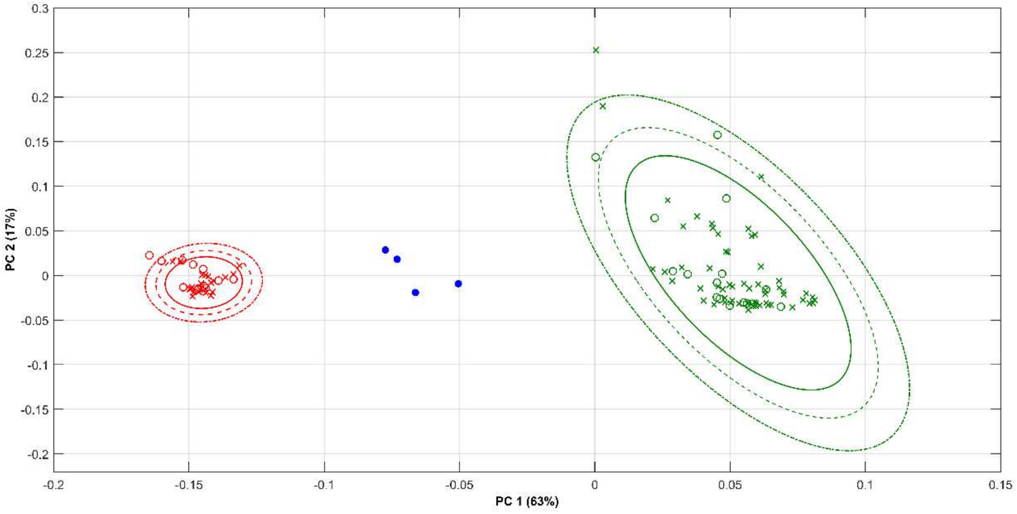

Figure 4-2: score plot of the two first PCs. Hardwood samples are marked in green, softwood in red and blends in blue. x: training set; ○: test set; ●: blend 50:50 (w/w %); solid line (-): 95%; dash line(- - -): 99%; dot line (•••): 99.9%.

The loading plot highlights which variables, and related compounds, mostly contribute at separating the hardwood and softwood samples (Figure 4-3). In Table 4-1 the most relevant wavenumbers associated to the differences between the two classes have been reported. A shift of wavenumbers can be noticed in Figure 4-3 with respect to Table 4-1 because of the use of first derivative as pre-treatment. Also other authors found that a similar band shift (i.e. between the position of the peak in the raw spectrum and in the derivative spectrum) in multivariate model development can be observed around the original peak position (Fagan, Everard et al. 2011, Zhou, Jiang et al. 2015, Zhou, Jiang et al. 2015). The interpretation of derivative spectra is in general more complex than raw spectra, in fact a peak of maximum absorbance on the original spectra corresponds to zero in the 1st derivative and to a trough in the 2nd derivative (Fagan, Everard et al. 2011). As consequence, in this study, zero points corresponding to derivative of the peak were selected in order to identify the compounds more influent for the PLS model.

The results are in line with literature reporting that the chemical differences between hardwood and softwood are related to lignin, hemicellulose and extractive compounds (Evans 1991, Faix 1991, Popescu, Singurel et al. 2009, Emandi, Ileana Vasiliu et al. 2011, Santoni, Callone et al. 2015). Lignin has different composition in softwood and in hardwood. In more detail, the softwood lignin is a polymer consisting mostly of guaiacyl units with only a small amount of p-hydroxyphenyl and syringyl units. The hardwood lignin is composed of syringyl and guaiacyl units with a small amount of p-hydroxyphenyl units (Santoni, Callone et al. 2015). Hemicelluloses are mainly composed by 80-90% of 4-O-methylglucoronoxylan in hardwoods (Sjöström 1993, Tumuluru, Sokhansanj et al. 2011), while 60-70% glucomannans and 15-30% arabinogalactan are the main constituents in softwoods (Grønli, Várhegyi et al. 2002, Lundqvist, Jacobs et al. 2003, Tumuluru, Sokhansanj et al. 2011).

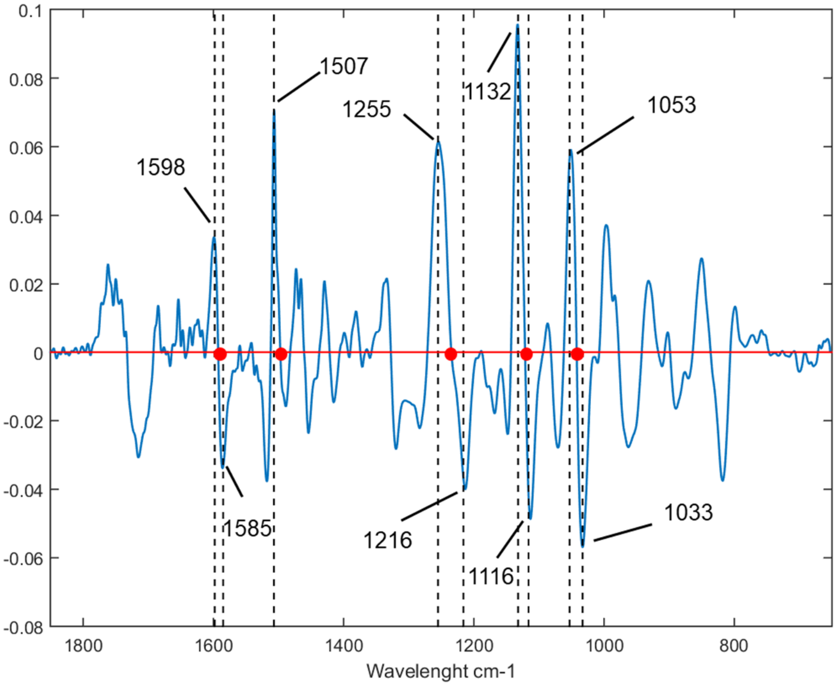

Figure 4-3: loading plot of PC1. The dotted lines refer to the spectral differences also reported in Table 4-1, the red dots refer to the wavenumbers searched in literature references for compounds assignment.

Table 4-1: main IR absorption bands associated for the discrimination between hardwood and softwood samples.

Measured wavenumber (cm^{-1})	**Bibliography wavenumber (cm^{-1})**	**Assignment**	**Compound**
1033 / 1053	- 1044 - 1045 - 1040	- arabinoxylans (Iroba and Tabil 2015) - C-OH bending (Iroba and Tabil 2015) - (s) C-C and (s) C-O (Popescu, Singurel *et al.* 2009)	- H - C, H - all
1116 / 1132	- 1122	- (b) C–H in plane in the syringyl monomer (Pandey and Pitman 2003)	- L
1216 / 1255	- 1240 - 1233 - 1226 - 1234, 1233, 1226-1234	- (s) C−O in the O=C−O group in xylose (Simonović, Stevanic et al. 2011) - syringyl nuclei of plane at positions 2, 5, and 6 in guaiacyl units (Zhou, Jiang et al. 2015) - syringil ring vibration; (s) C−O of lignin and xylan (Santoni, Callone et al. 2015) - syringyl ring breathing with (s) C-O, (s) C-C stretching, (b) OH in plane (Evans 1991, Popescu, Singurel *et al.* 2009) (REYES, TEIXEIRA MENDONÇA *et al.* 2013) (Emandi, Ileana Vasiliu et al. 2011)	- H - L - L, H - L, C

1507	- 1506–1513 - 1505 - 1510 - 1510 - 1515-1505 - 1510-1501	- (s) asymmetric vibration for aryl ring (Lupoi, Singh et al. 2015) - (s) guaiacyl and syringil (hardwood) (Esteves, Velez Marques et al. 2013) - (s) guaiacyl (softwood) (Esteves, Velez Marques et al. 2013) - specific band of aromatic skeletal vibrations(Popescu, Singurel *et al.* 2009) (Schwanninger, Rodrigues et al. 2004, Emandi, Ileana Vasiliu et al. 2011)	L
1585 -1598	- 1595 - 1596−1600 - 1594	- lignin absorption band (Bjarnestad and Dahlman 2002) - (s) symmetric vibration for aryl ring (Lupoi, Singh et al. 2015) - (s) C=C of the aromatic ring (syringyl), (s) C=O, and (b) C−H (Popescu, Singurel et al. 2009)	L

(s): stretching; (b): bending; L: lignin; H: hemicellulose; C: cellulose; E: extractives.

SIMCA computation is based on independent principal component analyses (PCA) carried out for each group in the training set to build the class models for hardwoods (n=60) and softwoods (n=22). A total of 10 PCs have been chosen for each class for a total explained variance of 96% and 94% for softwood and hardwood models, respectively. Considering a distance between the two models of 132.9, the separation between the two classes is significant and the model could be used for identification purposes. It should be taken into account that 3 is the limit to consider two classes significantly separated.

The discrimination power plot was investigated to identify the most relevant wavenumbers responsible in the separation between hardwood and softwood samples: higher the value, more influential the wavenumber is. The results confirm the interpretation of PCA loading. In fact, the wood compounds involved in separating the two classes are lignin (1116, 1132, 1216, 1255 and 1598 cm^{-1}) and hemicellulose (1216 and 1255 cm^{-1}), i.e. the biomass compounds that show different composition between hardwood and softwood.

In order to check the performance of the classification model, test set and blends samples were used for the external validation at 0.1% of significance level. The Si (sample-to-model distance) vs Hi (distance to the model center, leverage) plots (Figure 4-4) confirm the good performance of the developed models; 100% and 93% of correct classification for softwood and hardwood respectively. In fact, all the softwood samples were correctly identified and only one hardwood sample does not belong to its class because associated with a leverage value higher than the limit. As expected, the four blend samples were not classified in any class, suggesting the suitability of the model to discriminate between pure and blend samples. To take into account that only 50/50 (w/w %) blends of hardwood and softwood have been used in order to test the performance of the model with not pure samples. A further investigation might be to verify if the method would perform well on blends where one component is present at less than 50%.

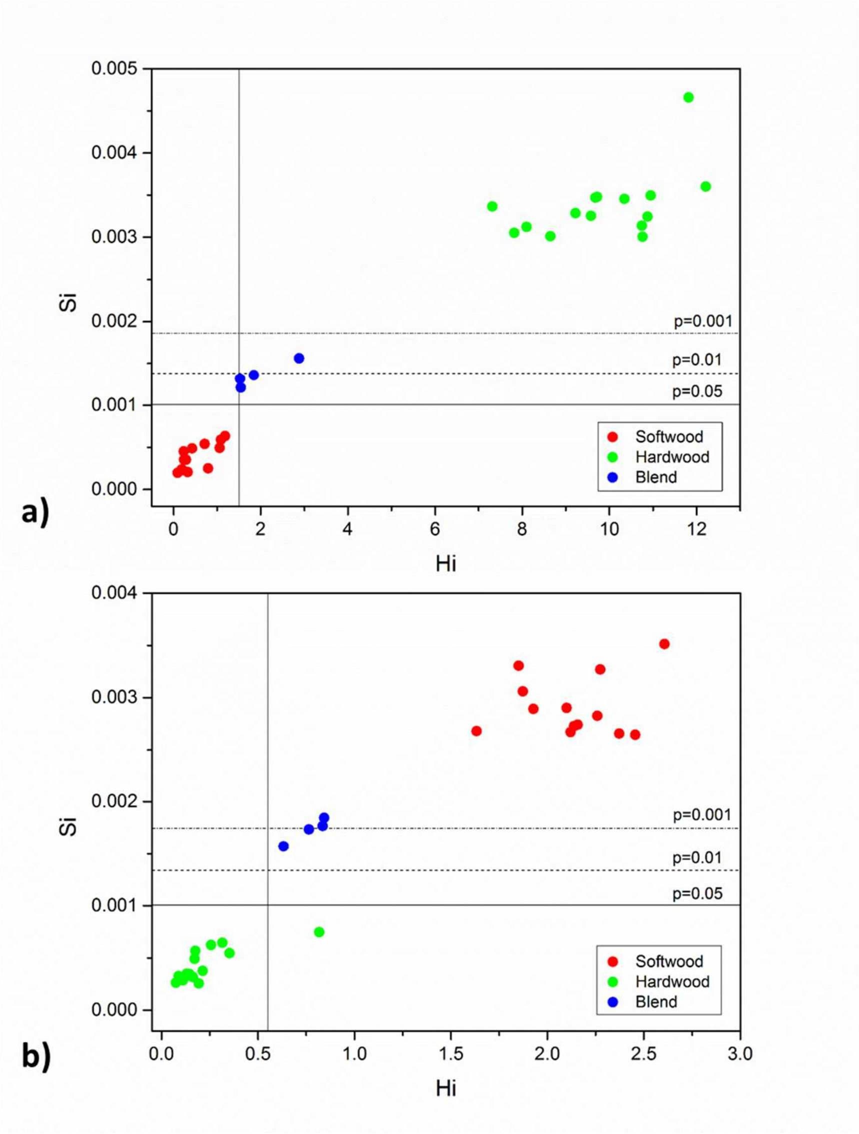

Figure 4-4: Si vs Hi plot for softwood (a) and hardwood (b) classes.

The results of the study showed that FT-IR coupled with SIMCA computation is an efficient tool for the discrimination of hardwoods, softwoods and their blends. This identification is particularly important since it allows to answer about the origin and source of the material, as requested by the technical standards EN ISO 17225, and could represent a decision-making tool for the different stakeholders. The operators involved in the pellet chain quality control could for sure benefit of such a method since it makes possible to recognize raw materials different from those declared and mixed during the pelletizing process.

Hardwood and softwood samples were also analysed using a FT-NIR spectrophotometer in order to verify the feasibility to use the near range of the infrared region (instead of the medium range) to get information about the origin and source of the material. In fact, in comparison to MIR spectroscopy, NIR spectroscopy has found many applications in industrial processes because of the higher energy level (in comparison to the fundamental vibrations of MIR), the higher sensitivity, the deeper penetration into organic samples and, most important, it is simpler to use, robust and accessible in terms of cost (Burns and Ciurczak 2008). The same workflow already described in this Chapter was followed for the samples preparation. A total of 4 species and 80 samples (20 samples for each species) were collected: fir and pine for softwood and beech and maple for hardwood. Maple species replaced the sessile oak mainly for two reasons. First of all to verify the possibility to discern this species from the others and secondly because most of the pellet on the European market is imported from United Stated and from Canada (Magelli, Boucher et al. 2009, Sun and Niquidet 2017, Eurostat Accessed: September 2017), where maple is a widespread species.

The spectra were pre-treated using second order derivatives (second polynomial order, 13 smoothing points with Savitzky-Golay (Savitzky and Golay 1964)) and the three replicates were averaged since no replicates outlier were found. Because of the low signal-to-noise-ratio, the last part of the spectrum (from 10,000 to 6,910 cm^{-1}) was removed and the PCA was computed on the spectral range between 6,910 and 4,000 cm^{-1}. Figure 4-5 shows the scores plot of PC1 versus PC3. PC1 and PC3 explained 58% and 3% of the total initial variability, respectively. The separation between the two classes is along PC1, whereas along PC3 it is possible to observe a separation between the different wood species – fir/pine and beech/maple.

The first loading was investigated to identify the main wavelengths responsible for the separation between softwood and hardwood samples (Figure 4-6). The wood compounds responsible for such separation are mainly hemicellulose (5809, 5242, 4640, 4490, 4440 and 4289 cm^{-1}) and lignin (5978 and 4440 cm^{-1}), confirming the results of the FT-IR analysis.

The study suggests the possibility to use both MIR and NIR spectroscopy technique for the discrimination between hardwood and softwood. In fact, as the separation between the two classes is so clear, a classification model could be developed using FT-NIR spectra and surely, it would perform well. Furthermore, the study has demonstrated that more detailed information at the qualitative level about wood species can be obtained. In summary, infrared spectroscopy represents a first step in intensive monitoring of the product, addressing the quality assurance requirements (EN ISO 17225).

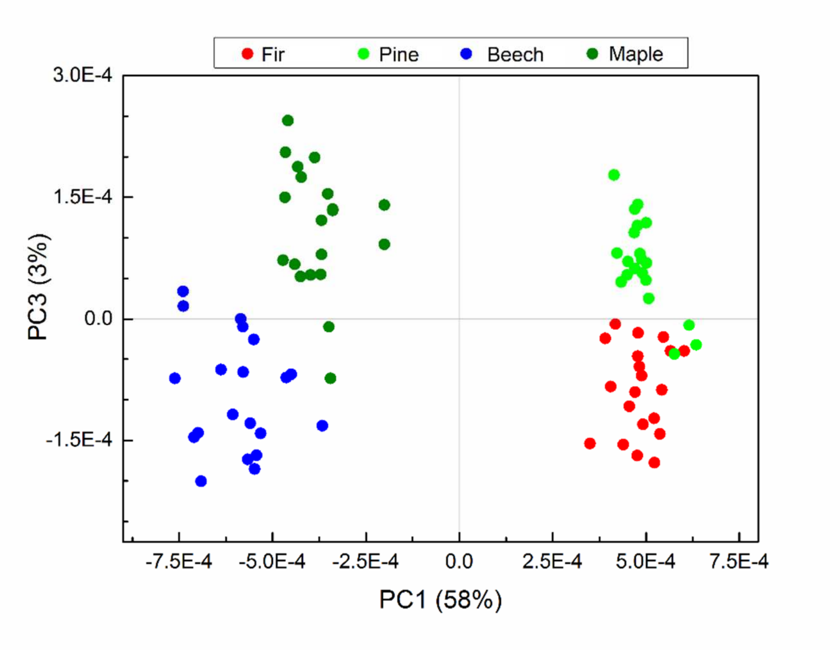

Figure 4-5: score plot of PC1 versus PC3.

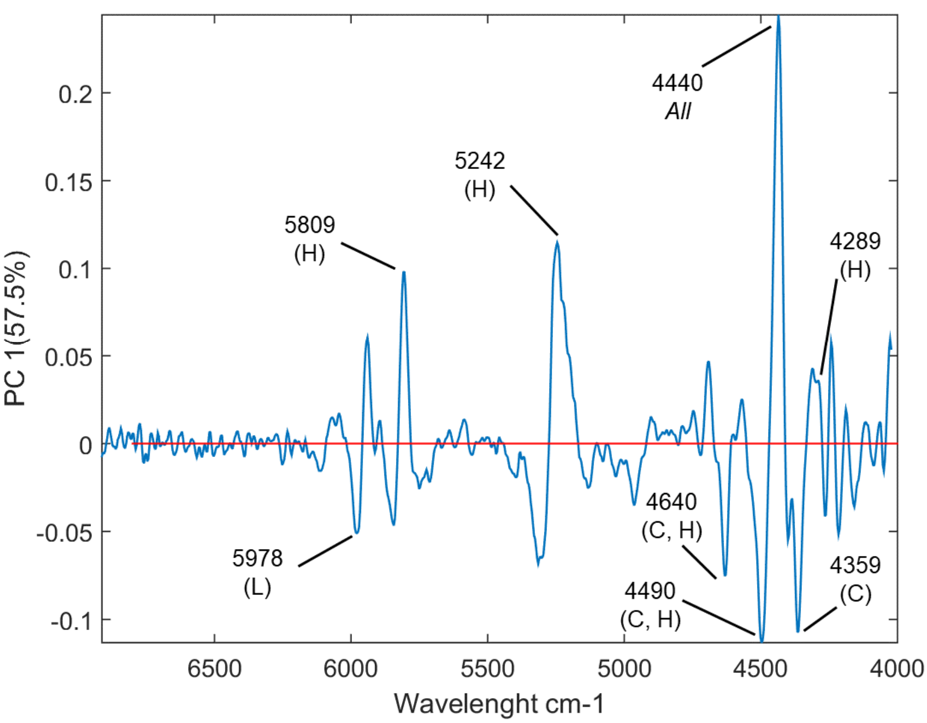

Figure 4-6: loading plot of PC1 (L: lignin; H: hemicellulose; C: cellulose).

4.2 Prediction of softwood/hardwood contents in blends *(Paper II)*

This section is mainly based on the paper "Prediction of Hardwood and Softwood Contents in Blends of Wood Powders Using Mid-Infrared Spectroscopy" accepted for publication in the peer-reviewed journal *Energy & Fuels*.

The good ability of the PCA model of **Paper** I in separating hardwood and softwood samples and the location of the four blends right in the middle of the score plot, showing the right proportionality of hardwood/softwood content (§ Figure 4-2, Chapter 4.1), suggests the suitability of the method in identifying the blends percentage. As consequence, the purpose of the presented study is the prediction of hardwood and softwood contents in blend samples by using FT-IR analysis coupled with PLS computation.

Fir (softwood), pine (softwood), sessile oak (hardwood) and beech (hardwood) samples, some of the most common European species in the energy pellet sector, have been selected and collected. Subsequently, 29 hardwoods and 32 softwoods have been used for the preparation of 28 blends at different concentration levels. The blending degrees were between 5 and 95 w/w % of hardwood/softwood content, in particular the following concentrations were used 5, 15, 25, 50, 75, 85, 95 w/w %. The workflow followed for the preparation of the samples and of the blends is the same as described in **Paper** I (§ Chapter 4.1). All the samples were analysed in medium infrared using a FT-IR spectrophotometer.

First of all PCA was computed on the pre-treated spectral dataset in order to explore the overall statistical variance, search for clustering among the samples and recognize blends at different levels of concentration. The *fingerprint region* from 1850 cm^{-1} up to 650 cm^{-1} was selected and all the spectra were subjected to different pre-treatments. SNV and 1st derivative (second polynomial order, 13 smoothing points with Savitzky-Golay (Savitzky and Golay 1964)) results to be the best pre-treatment. The three replicates of each sample were averaged since no replicate outlier was found.

The PCA results confirm those of **Paper** I (§ Chapter 4.1). In fact, the score plot (Figure 4-7) shows that softwoods have negative PC1 scores, whereas hardwoods have positive PC1 scores. Moreover, along PC2 samples are separated according to the species: pine and sessile oak have mainly positive PC2 scores and beech and fir have negative PC2 scores. The identification is based on the chemical differences between hardwood and softwood with regards to lignin, hemicellulose and extractives fractions.

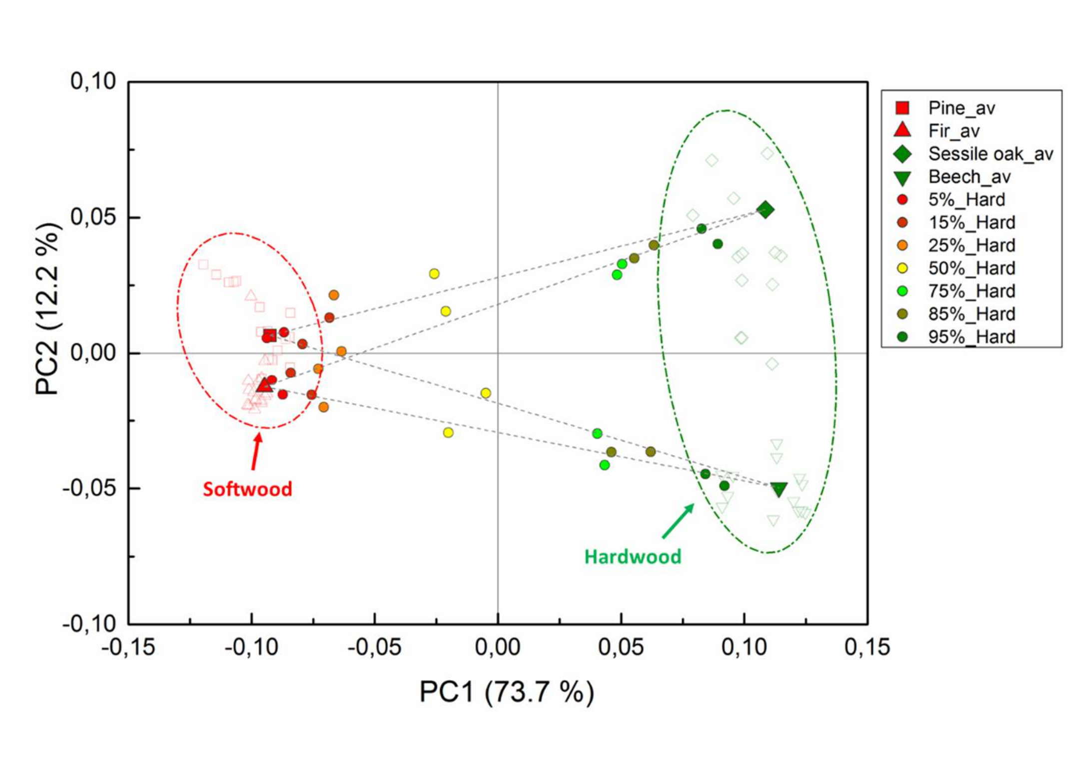

Figure 4-7: PCA score plot of hardwood, softwood and blends samples at different levels of concentration. Species_av: average samples, mean of the 10 pooled samples for each species; %_Hard: percentage of hardwood in blend; □: pine samples; Δ: fir samples; ◊: sessile oak samples; ∇: beech samples.

The linear trend of 50/50 (w/w %) blends of hardwood and softwood highlighted in **Paper** I is now much more evident, confirming that building a regression model with the purpose of estimating the amount of hardwood and softwood contents in blends is possible. The best PLS model for the prediction of hardwood and softwood contents was developed on the *fingerprint region* from 1,850 cm^{-1} up to 650 cm^{-1} using SNV and first derivative (Savitzky-Golay method, 13 smoothing points) as pre-treatments. As no replicate outlier was found, the three replicates of each sample were averaged.

The prediction model was developed considering 10 LVs and was validated using Leave-One-out Cross validation. The reference parameter is the softwood/hardwood content in blend samples. The accuracy was evaluated using the figures of merit, in particular coefficient of determination (Q^2 when it is calculated on predicted or cross-validated values and R^2 when it is calculated on calibration values) and RMSECV. The coefficient of determination (Q^2) measures how well the predicted values agree with the reference data: values between 0.50

and 0.65 allow the discrimination between high and low concentrations, between 0.66 and 0.81 indicate approximate quantitative predictions, between 0.82 and 0.90 reveal good predictions and models with Q^2 above 0.91 are considered to be excellent (Williams 2001, Fagan, Everard et al. 2011). The ratio of standard error of prediction to standard deviation (RPD) and range to error ratio (RER) were also calculated to assess the prediction accuracy of the model. The former is calculated by dividing the standard deviation of the response parameter by the RMSECV, while the latter as the ratio between the range of the response parameter and the RMSECV. According to Williams (Williams 2001), for screening application RPD value should be greater than 3, while RPD values greater than 5 can be used for quality control. RER values of less than 6 indicate a poor model not recommended in any application; while values between 7 and 20 classify the model as fair and may be used in screening applications; values greater than 20 indicate good predictions suggesting that the model could be suitable in any quality control application (Williams 2001, Fagan, Everard et al. 2007, Fagan, Everard et al. 2011). The developed model has RMSE, 0.22%, RMSECV, 3.78%; R^2, 1.00, Q^2, 0.99; RER, 26.4 and RPD, 9.5. All values indicate that the statistical method results to be excellent and the model could be used in any quality control application.

The first three loadings were analysed in order to identify the compounds associated to the most important wavenumbers linked to blend percentage prediction (Figure 4-8). Once again, lignin and hemicellulose are the most relevant components affecting the discrimination model because of their different composition in hardwood and softwood (§ Chapter 4.1). In detail, the main wood compounds that contribute to PLS prediction model are hemicellulose (1188, 1244, 1454 cm^{-1}), lignin (1068, 1143, 1244, 1270, 1420, 1512 cm^{-1}), cellulose (1420 cm^{-1}) and extractives (1697 cm^{-1}).

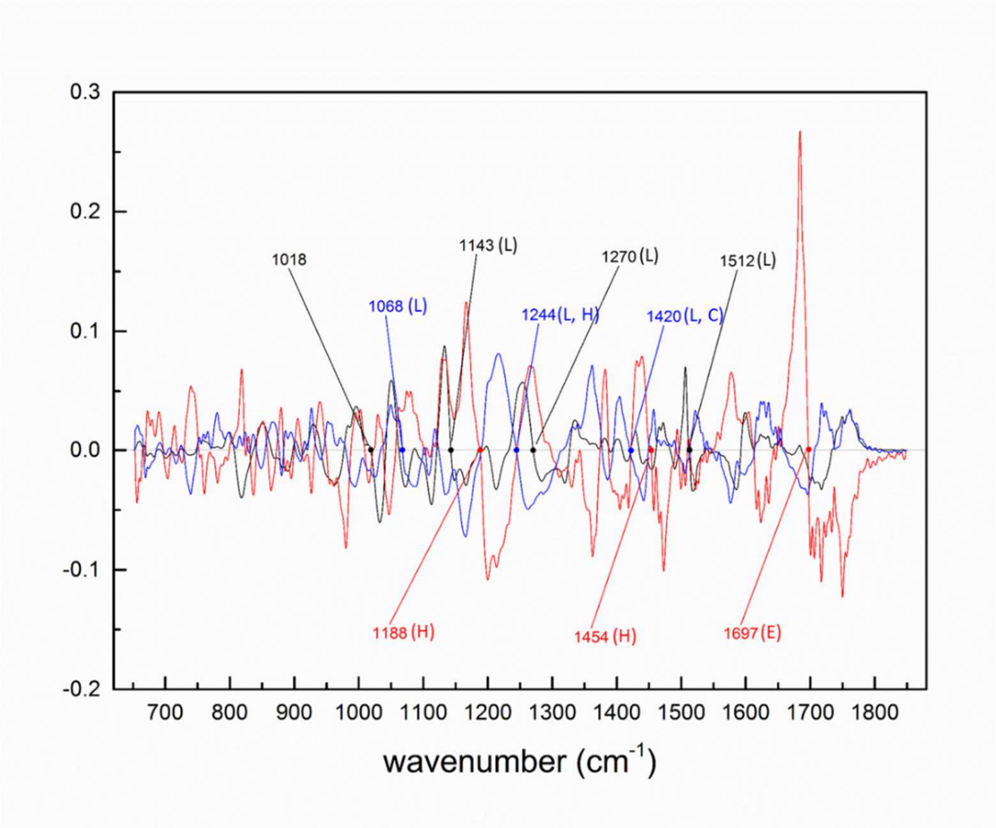

Figure 4-8: first three loading for the prediction of hardwood and softwood contents in blends samples (L: lignin; H: hemicellulose; C: cellulose; E: extractives).

The number of LVs was chosen considering the lower RMSECV value, but this can implicate an overfitting of the model and the difference between RMSE and RMSECV values could represent an evidence of it. This means that the model undergoes to minor fluctuations in the training data and it fits perfectly with the data of this study, but the performance would decrease trying to predict an unknown sample (it is like a tailored suit on a man!). The loading plot, the regression coefficient plot and the permutation test results were taken into consideration in order to choose the right number of LVs, trying to avoid overfitting and to model noise. It results that the spectral information contained in the data matrix could be explained considering only 4 LVs (Figure 4-9). RMSE increased from 0.22% to 3.04% and RMSECV from 3.78% to 4.26%, instead Q^2 remained unchanged (0.99), RER value is 23.5 and RPD 8.5. Even if the performance of the model could seem worst, the predictions result better.

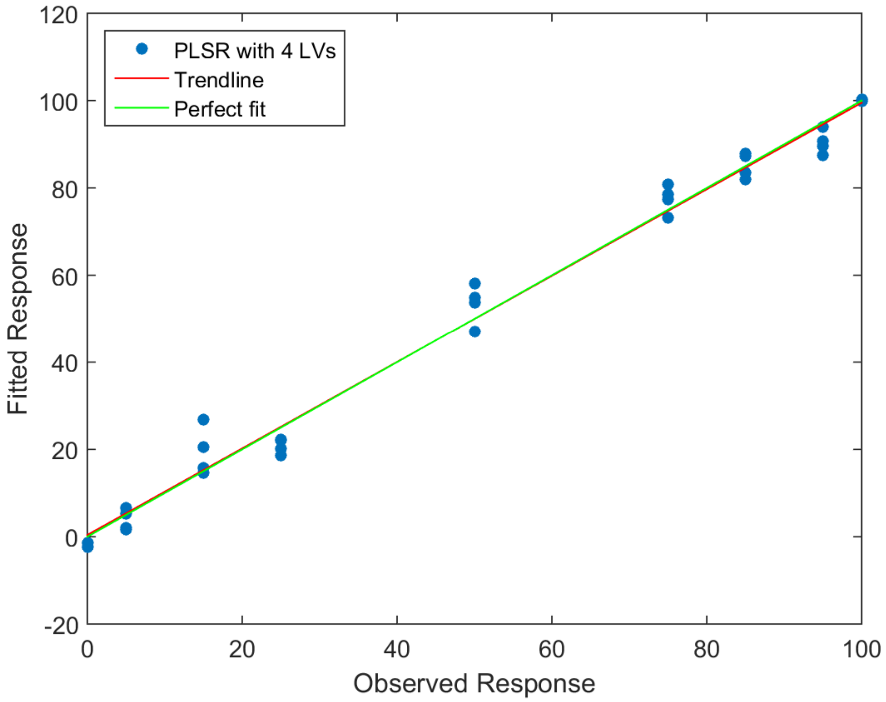

Figure 4-9: PLS regression model for the prediction of hardwood and softwood contents in blends samples.

The results of this study demonstrate the feasibility of IR spectroscopy analytical tool for the prediction of hardwood and softwood contents in blends. Moreover, the PCA results suggest that a discrimination model for the four species considered in the study (fir, pine, sessile oak and beech) would perform well. The prediction of hardwood/softwood content in blends will open up different scenarios for the sector stakeholders. First of all, the method could be useful at the production level in providing information about biofuel traceability, process tuning and quality control. Furthermore, it should engage attention of the producers and distributors of the biofuels and the operators involved in the quality control of the market. In fact, such a tool would permit to check for non-conformity and verify the compliance to product labels, especially pellet.

4.3 Discrimination between bark and wood *(Paper III)*

This section is based on the paper "The Use of Near-Infrared (NIR) Spectroscopy and Principal Component Analysis (PCA) To Discriminate Bark and Wood of the Most Common Species of the Pellet Sector" accepted for publication in the peer-reviewed journal *Energy & Fuels*.

The presence of bark affects the biomass quality, influencing mainly the ash content, and the standards EN ISO 17225 classify on the basis of this parameter, too. The determination of the bark presence using infrared spectroscopy would allow biomass producers and suppliers to guarantee that the products fulfil the requirements for the biomass quality, especially on milled or densified biomass for which is hard to determine. Furthermore, there are no good standardized methods to identify and determine the presence of bark (Harkin and Rowe 1971) and, at present, it is only declared by the producers with few practical possibilities to verify these claims.

Some of the most common European species in the energy pellet sector, i.e. fir, pine and beech, have been considered and selected for this study. Whole pieces of wood as tree log disks wood slices, beams or boards taken from different sawmills have been collected to obtain the wood samples dataset (n = 60). After 24 hours stabilization at 45 °C, the material reached the same moisture content of 6-7% and bark samples (n = 32) can be obtained by manually debarking wood slices (Figure 4-10). All the samples were first chopped into smaller pieces, and subsequently ground to below 1 mm of particle size by means of a cutting mill.

Successively, the same quantity of material of 10 pooled samples of bark and wood for each species has been used to prepare six average samples, i.e. wood and bark of beech, pine and fir. All the blends are composed by mixing the bark and wood average samples of the same species, obtaining a total of 30 blends at different bark concentration levels. The blending degrees were between 2 and 20 w/w % of bark in 2 w/w % increments. FT-NIR spectrophotometer was used to collect the near infrared spectra of all the samples and blends.

Figure 4-10: operation of manually debarking of tree log disks wood slices.

The pre-treatment of spectra has consisted in performing the second order derivative to enhance spectral differences and remove both additive and slope effects in the spectra (Rinnan, Berg et al. 2009). Savitzky-Golay (Savitzky and Golay 1964) algorithm with 13 smoothing points and second order polynomial was used to smooth the derivative spectra. Successively, since no replicate outlier was found and in order to reduce the spectra variation, an average of the two single measurements on the same sample has been calculated. Moreover, as the most significant information is contained between 7,500 and 4,000 cm^{-1}, only this part of the spectra was considered for interpretation and exploratory purposes.

Figure 4-11 and Figure 4-12 are used as support for interpreting the FT-NIR measurements and their links to the related wood compounds. In detail, the most important differences in the curves of Figure 4-11 explained the dissimilarities in chemical composition between bark and wood. In bark, lignin concentration is higher than in wood (Harkin and Rowe 1971), as well as the concentration of inorganic material (Harkin and Rowe 1971, Miranda, Gominho et al. 2012). Also extractives are both much more abundant and more variable in bark than they are in wood, with many types being present only in the bark (Harkin and Rowe 1971, Miranda, Gominho et al. 2012). Moreover, bark has substantially less polysaccharides than wood (Harkin and Rowe 1971).

Referring to Figure 4-11, the peak 1 (7012 cm^{-1}), related to the amorphous region in cellulose, shows some small differences between bark and wood (Sandak, Sandak et al. 2010). Minimal differences are reported in the range between 2 (6291 cm^{-1}) and 6 (5469 cm^{-1}). No differences are noticed in the region of cellulose between 6800 cm^{-1} and 6350 cm^{-1}, demonstrating that this area is minimally affected by the variations between bark and wood. Minute differences are observed in the peak 7 (5219 cm^{-1}). This band is the characteristic band of water, related to the combination of O-H stretching and O-H deformation vibration (Sandak, Sandak et al. 2010), proving that the moisture content between bark and wood samples is similar and did not affect the discrimination. The range from 10 (4890 cm-1) to 11 (4620 cm^{-1}) also shows some differences. These bands are normally assigned to O-H stretching + C-H deformation in cellulose (Sandak, Sandak et al. 2010).

Finally, evident differences are noticed in the peaks 13 (4405 cm^{-1}) and 16 (4281 cm^{-1}). These bands are assigned to semi-crystalline and/or crystalline regions in cellulose, in particular to C-H stretching + C-H deformation and C-H_2 stretching + C-H_2 deformation, respectively (Sandak, Sandak et al. 2010, Schwanninger, Rodrigues et al. 2011). Peaks related to cellulose – 14 (4358 cm^{-1}), 17 (4252 cm^{-1}) and 19 (4158 cm^{-1}) – and holocellulose – 9 (4964 cm^{-1}), 12 (4493 cm^{-1}) and 18 (4196 cm^{-1}) – are confirming the differences between bark and wood and indicate that the amount of polysaccharides is higher in wood than in bark.

The spectral differences between hardwoods and softwoods (Figure 4-12) are linked to hemicellulose and lignin components, in line with literature (Evans 1991, Faix 1991, Popescu, Singurel et al. 2009, Emandi, Ileana Vasiliu et al. 2011, Santoni, Callone et al. 2015) and with the results of **Paper** I and **Paper** II (§ Chapters 4.1 and 4.2).

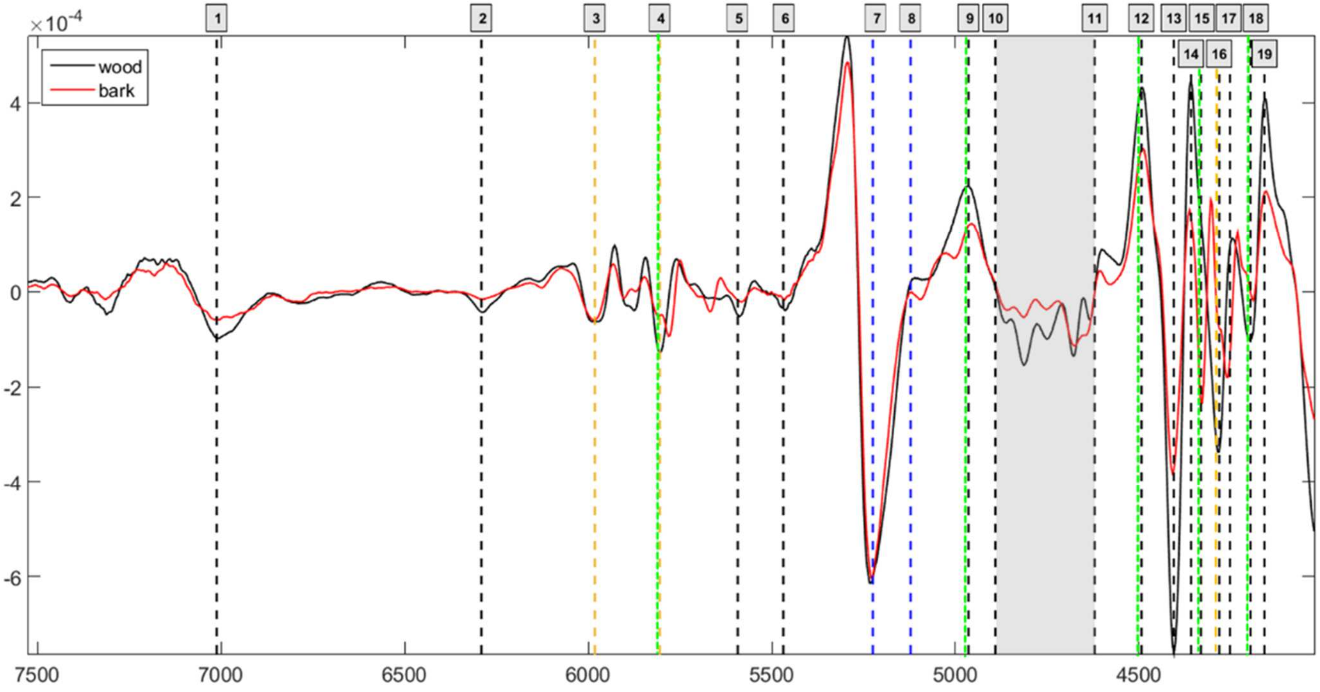

Figure 4-11: spectra pre-treated with second derivatives of bark samples (red line) and wood samples (black line). Dotted lines marked with numbers refer to the main spectral differences, in detail green line: hemicellulose; black line: cellulose; orange line: lignin, blue line: water.

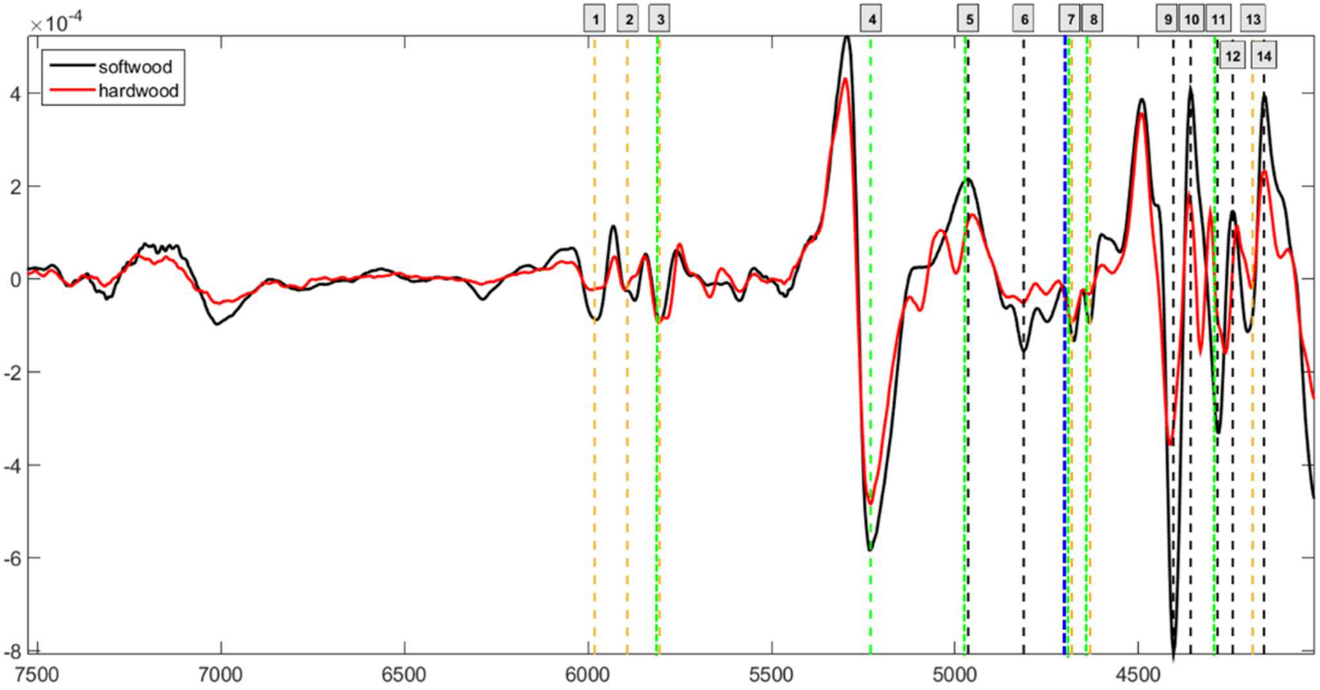

Figure 4-12: spectra pre-treated with second derivatives of hardwood samples (red line) and softwood samples (black line). Dotted lines marked with numbers refer to the main spectral differences, in detail green line: hemicellulose; black line: cellulose; orange line: lignin, blue line: extractive.

PCA computation was performed on the pre-treated data matrix of samples and on mean centered data in order to explore the information about biomass origin and source from spectral data. Figure 4-13 shows the PCA score plot of the two first PCs. The wood and bark samples were well separated along PC1: bark samples have negative PC1 scores whereas wood samples have positive PC1 scores. In addition, it is possible to observe a separation between hardwood and softwood along PC2, both for wood and bark. Because of the high heterogeneity of bark samples, this separation is more evident for wood samples than for bark samples (Harkin and Rowe 1971). In this regard it is possible to see that a beech bark sample falls into the softwood bark grouping, confirming the high variability of the bark, especially that of beech.

A confidence ellipse was calculated at a specific level of significance (99% confidence level) both for bark, hardwood and softwood groups. As only a sample of fir, a sample of pine and a sample of beech appear outside the 99% confidence limits (ellipsoids), PCA could be considered a powerful tool not only for differentiating between bark and wood samples, but also between hardwoods and softwoods.

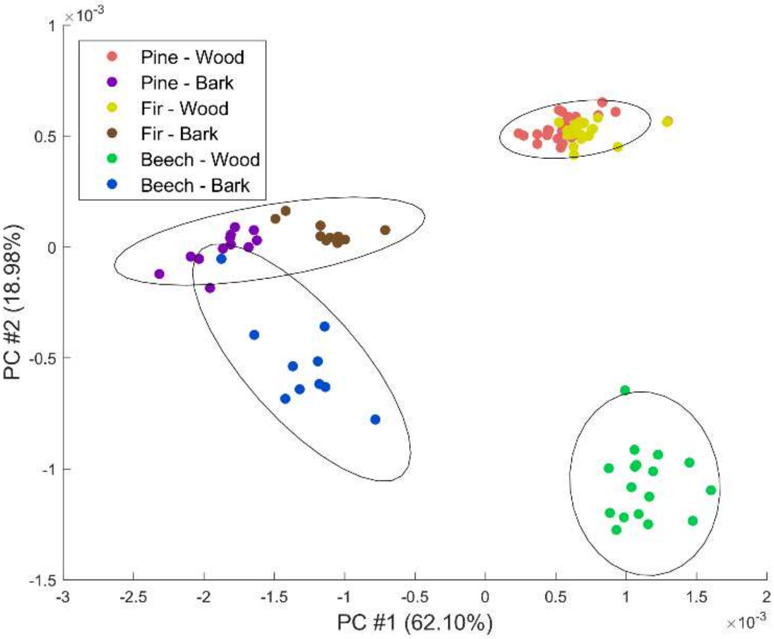

Figure 4-13: score plot of the two first PCs.

The first loading was investigated to figure out which variables – and the related compounds (lignin, L; cellulose, C; hemicellulose, H; extractives, E) – are associated to the discrimination between bark and wood, whereas the second loading was used to investigate the variables responsible for the discrimination between hardwood and softwood samples. The interpretation of the loadings might be explained with support of Figure 4-11 and Figure 4-12, where the characteristic spectral bands have been already detected.

Referring to Figure 4-14.a, it results that bands 1 (7012 cm^{-1}), 2 (6291 cm^{-1}), 3 (5982 cm^{-1}), 5 (5593 cm^{-1}), 6 (5469 cm^{-1}), 7 (5219 cm^{-1}) and 8 (5122 cm^{-1}) have a minimal impact on the discrimination between bark and wood samples. Instead bands 4 (5805 cm^{-1}; H, L), 9 (4964 cm^{-1}; C, H), 12 (4493 cm^{-1}; C, H), 14 (4358 cm^{-1}; C), 13 (4405 cm^{-1}; C), 15 (4331 cm^{-1}; C, H), 16 (4281 cm^{-1}; C), 17 (4252 cm^{-1}; C), 18 (4196 cm^{-1}; C, H) and 19 (4158 cm^{-1}; C), show high peaks in the loading plot, meaning that these bands are important for the discrimination of the two clusters. Also the range between 10 (4890 cm^{-1}) and 11 (4620 cm^{-1}) seems to be important, especially the two bands at 4651cm^{-1} and 4636 cm^{-1}.

Referring to Figure 4-14.b, bands 1 (5982 cm^{-1}; L), 3 (5805 cm^{-1}; H, L), 4 (5230 cm^{-1}; H), 7 (4682 cm^{-1}; L, E, H), 8 (4632 cm^{-1}; L), 10 (4358 cm^{-1}; C), 9 (4405 cm^{-1}; C) and 12 (4243 cm^{-1}; L) are associated with high peaks, meaning that these peaks are discriminative for the difference between the two clusters, while bands 2 (5893 cm^{-1}), 5 (4964 cm^{-1}), 6 (4813 cm^{-1}), 11 (4285 cm^{-1}), 13 (4189 cm^{-1}) and 14 (4158 cm^{-1}) have minimal impact. The results are in line with those obtained in **Paper** I and in **Paper** II (§ Chapters 4.1 and 4.2).

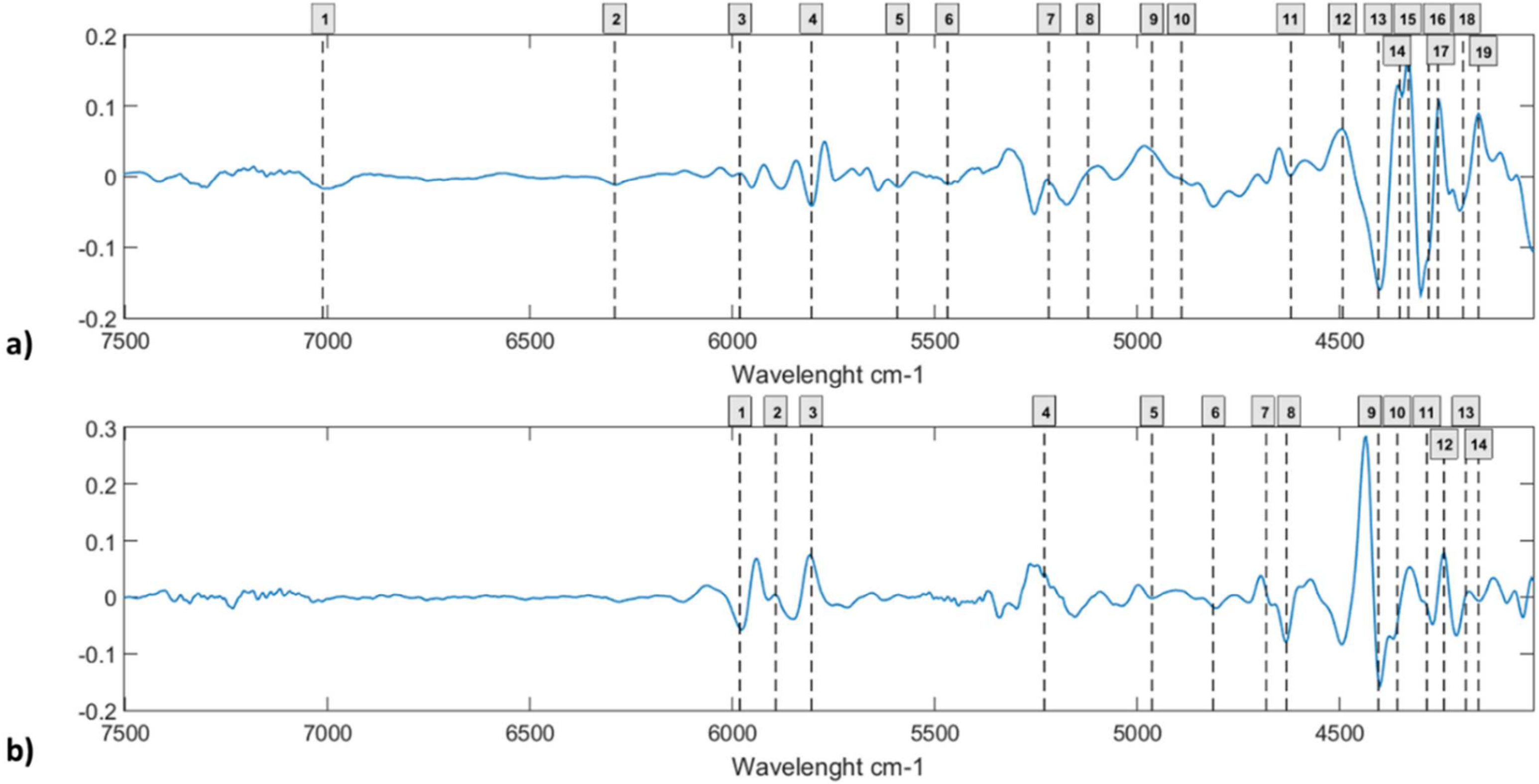

Figure 4-14: first loading, with the important wavenumbers for the discrimination between wood and bark marked with dotted lines (a). Second loading, with the important wavenumbers for the discrimination between hardwood and softwood marked with dotted lines (b).

In order to inspect the performance of the PCA model, the calculated loadings were used to project the spectra of blends between wood and bark (ten level of concentrations from 2 to 20 w/w %) in the PCA score plot (Figure 4-15). The different blends are positioned mainly between wood and bark clusters along the lines of connection of the average species samples showing a linear trend. It is also evident that blends with high bark concentration are well separated both from wood and bark groupings in agreement with their known percentage composition. Oppositely, those with low bark concentration fall in the wood grouping.

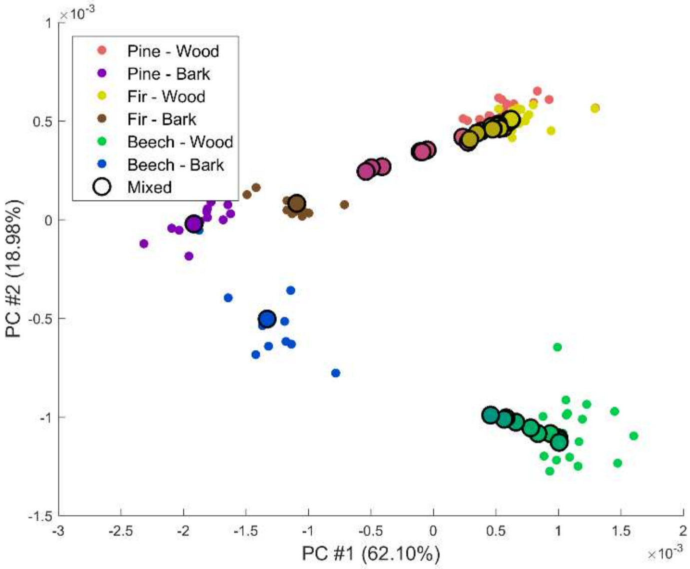

Figure 4-15: projection of blends between wood and bark at different levels of concentration in the PCA score plot.

EN ISO 17225-2 (EN ISO 17225-2:2014) is the quality standard for graded wood pellets and divides the product in three quality classes (A1, A2 and B) according to different attributes and chemical-physical parameters. The most important discriminant parameter is the ash content. The limit between A1 and A2 quality classes is 0.7% of ash content. According to literature, debarked wood has at around 0.3% of ash content (Filbakk, Jirjis et al. 2011, Tooyserkani, Sokhansanj et al. 2013, Simon, Brostaux et al. 2014, Yildiz, Ronsse et al. 2015), while bark contains around 2% (Lehtikangas 2001, Filbakk, Jirjis et al. 2011, Tooyserkani, Sokhansanj et al. 2013, Feng, Yuan et al. 2014, Feng, Yuan et al.

2016), meaning that a bark content up to 20% could be added without changing the quality classification (A1).

It is therefore clear the great potential of the developed model. Since it is able to recognize 20% bark content in blend samples at 99% confidence interval, it allows checking the pellet quality and recognizing possible blends, avoiding problems in terms of ash quantity and emissions. Such a tool could be used by the sector operators that want to check the product compliance or be applied directly in the production line. In addition, building a regression model (PLS) with the purpose of estimating the amount of bark in wood samples seems highly likely, given the linear trend in the blends along the axis of pure wood to pure bark. Lastly, the evident separation between the different classes in the PCA score plot suggests that the development of a discrimination model could be a further perspective.

4.4 Discrimination between treated and virgin wood *(Paper V* and *Paper VI)*

This chapter presents the results of near infrared measurements and multivariate statistical modelling for the discrimination between virgin wood and residues from wood processing industry (hereafter referred as treated wood). The purpose is to verify if wood-based material is significantly different from virgin wood and, if it is, to develop a model for unknown samples classification. This is particular relevant because in accordance to EN ISO 17225-2 (EN ISO 17225-2:2014), pellet B could be made up by by-products and residues from wood processing industry, but the use of these types of material for pellet production is not allowed in some EU countries. For instance, Italian national legislation (D.lgs. n. 152/2006) considers these residues as waste. As consequence, the development of a rapid and cheap tool based on near infrared analysis is essential for the quality control of the product. This section is based on the results presented during the 18th International Conference on Near Infrared Spectroscopy in Copenhagen (**Paper** V) and the submitted paper "Near infrared spectroscopy for the discrimination between different residues of the wood processing industry in the pellet sector" (**Paper** VI).

A set of 106 wood samples including virgin wood and four different types of residues from wood processing industry - glue-laminated wood, plywood, oriented strand board (OSB) and chipboard - have been used in this study (Figure

4-16). Beams, boards and whole pieces of wood as tree log disks wood slices were collected in different sawmills and they could be considered representative of the national scene, since most of the wood-based materials are in any case imported from abroad. Only fir, pine and beech – three of the most important species of the pellet sector – were included in the virgin wood dataset.

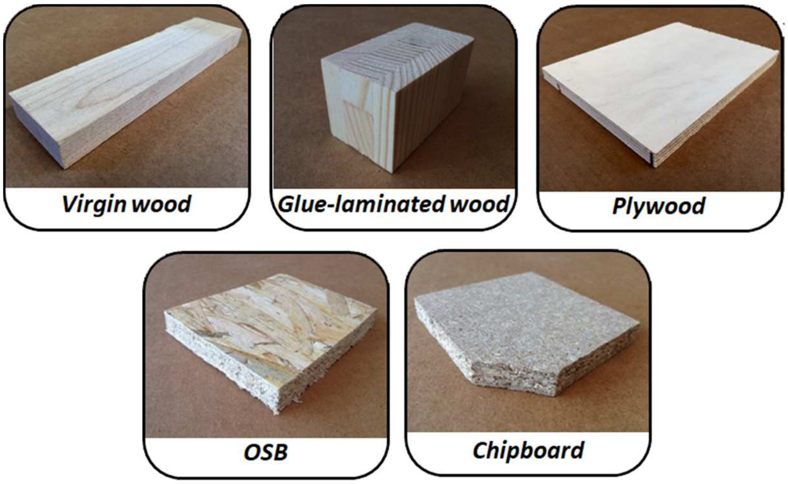

Figure 4-16: wood samples considered in this study; virgin wood and four different types of residues from wood processing industry - glue-laminated wood, plywood, oriented strand board (OSB) and chipboard.

First of all, the wood slices were conditioned for 24 h at 45 °C in order to limit the moisture content variability among the samples. Later, all the samples underwent to a reduction in smaller pieces and grinding below 1 mm of particle size by means of a cutting mill. At the end, they were analysed by near infrared spectroscopy.

Two types of glues, i.e. vinyl and polyurethane glue, have been also collected and analysed after drying and polymerization process. To take into consideration that the resin-binders normally used in the wood processing industry are phenol-formaldehyde, urea-formaldehyde, melamine-formaldehyde, and isocyanate (Stark, Cai et al. 2010) and the consolidation of the system wood particles-adhesives is normally due to heat and/or pressure process (Stark, Cai et al. 2010), changing the structure of the final product. In any case, considering the difficulties in finding the regular industrial glues and in reproducing the bonding conditions, the two glues were analysed only for interpretation purposes.

All the spectra were pre-processed using the second order derivative (Savitzky-Golay (Savitzky and Golay 1964) with 21 smoothing points and 2nd order polynomial) to minimize the effect of baseline shifts and noise. Subsequently, PCA computation was performed in order to search for clusters in the samples and assess the possibility to develop a successive classification model (i.e. PLS-DA). Looking at the PCA score plot (Figure 4-17), some clustering (or separation trends) are clear. Chipboard, OSB and plywood samples are well grouped and separated from virgin wood, but OSB is also scattered in the score plot. In fact, the great variability of applications for which it could be used made this material really heterogeneous. The key-element is the type of glue, different on the basis of the applications (internal or structural). Moreover, the location of OSB in the plot is also influenced by the presence of hardwood and/or softwood strands making up the material.

Glue-laminated wood is not so well separated from virgin wood, but both groups are well located along PC1 in accordance to hardwood or softwood composition: hardwoods have negative PC1 scores whereas softwoods have positive PC1 scores.

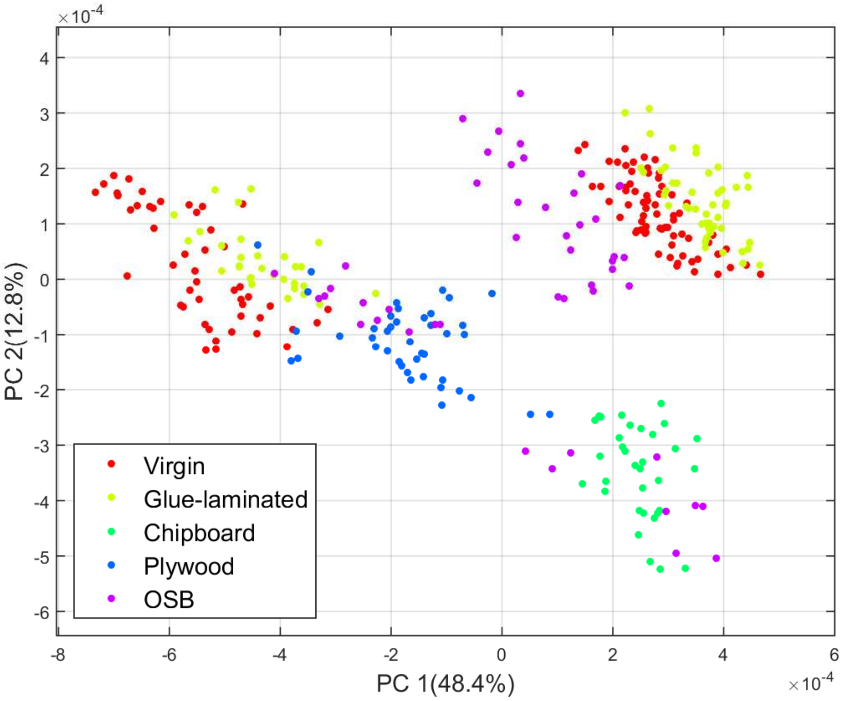

Figure 4-17: score plot of the two first PCs.

The compounds responsible for the separation between treated and virgin wood samples were detected by investigating the loading plot (Figure 4-18) and the spectral differences in the FT-NIR measurements. It results that the discrimination between the two clusters is mainly due to the N- bonds, present only in the glued-wood, and to the amount of C-H bonds of the material, as the compounds of wood (i.e. lignin, cellulose and hemicellulose) have a higher number of these bonds in comparison to the treated wood (Dessipri, Minopoulou et al. 2003, Minopoulou, Dessipri et al. 2003, Tomlinson, Ghita et al. 2006). Among all the wavenumbers detected, the most relevant were at 4440 cm^{-1} related to the CH_2 combination of methylol group (Dessipri, Minopoulou et al. 2003, Minopoulou, Dessipri et al. 2003), 5242 cm^{-1} assigned to 1st overtone of -CN functional group (Tomlinson, Ghita et al. 2006), 5311 cm^{-1} related to C–H overtone of aldehyde groups (Dwivedi and Rai 2009) and 5805 cm^{-1} assigned to 1st overtone C–H stretching of methyl and methylene structures of glue (Minopoulou, Dessipri et al. 2003, Tomlinson, Ghita et al. 2006, Dwivedi and Rai 2009).

Bands at 5331 (Yan, Villarreal et al. 2013), 4910 (Popescu and Popescu 2013), 4505 (Pecoraro, Pizzo et al. 2015), 4382 (Schwanninger, Rodrigues et al. 2011), 4359 (Schwanninger, Rodrigues et al. 2011) and 4324 (Pecoraro, Pizzo et al. 2015) cm^{-1} are mostly linked to C– bonds, more abundant in the compounds of wood than they are in the treated wood. A more accurate description of the detected wavenumbers responsible for treated/virgin wood separation is reported in **Paper** V appended to this thesis.

The most relevant wavelengths assigned to glue compounds were compared with the spectra of vinyl and polyurethane glue. Figure 4-19 reports the mean spectrum of virgin wood samples and the spectrum of vinyl and polyurethane glue and confirms the interpretation performed. Indeed, almost all the wavelengths assigned to glue correspond to peaks in the vinyl and polyurethane spectra. Some differences can be noticed; in fact some wavelengths assigned to wood correspond also to peaks in the two glue spectra. This is due to the different types of glue and bonding conditions in comparison to those normally used in the wood processing industry.

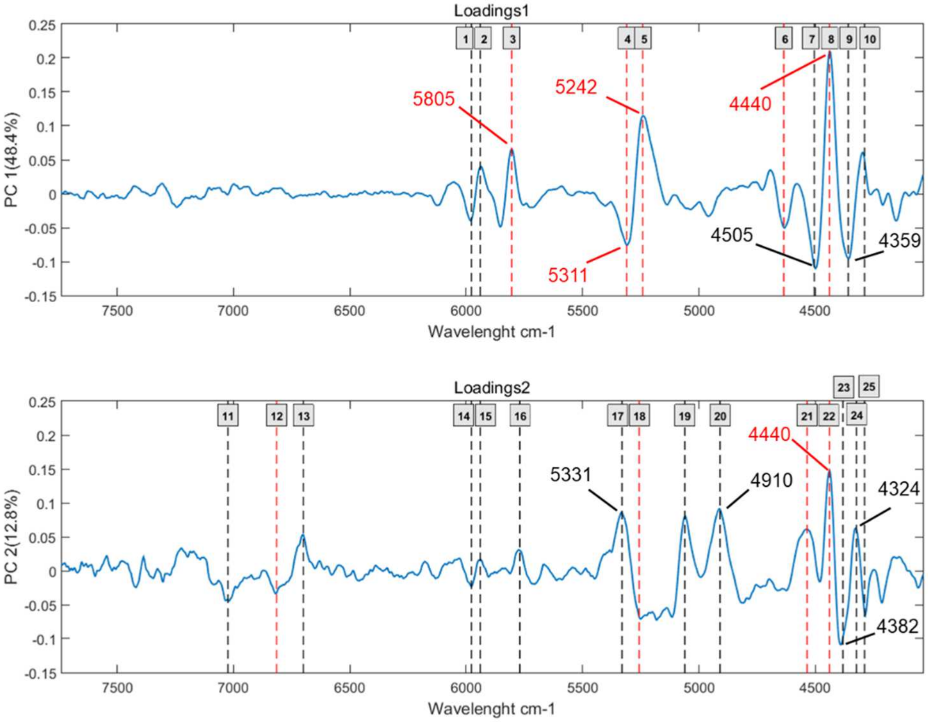

Figure 4-18: first two PCA loadings. Important wavenumbers for the discrimination between treated and untreated wood samples are marked with red (wavelengths assigned to glue) and black (wavelengths assigned to wood) dotted lines.

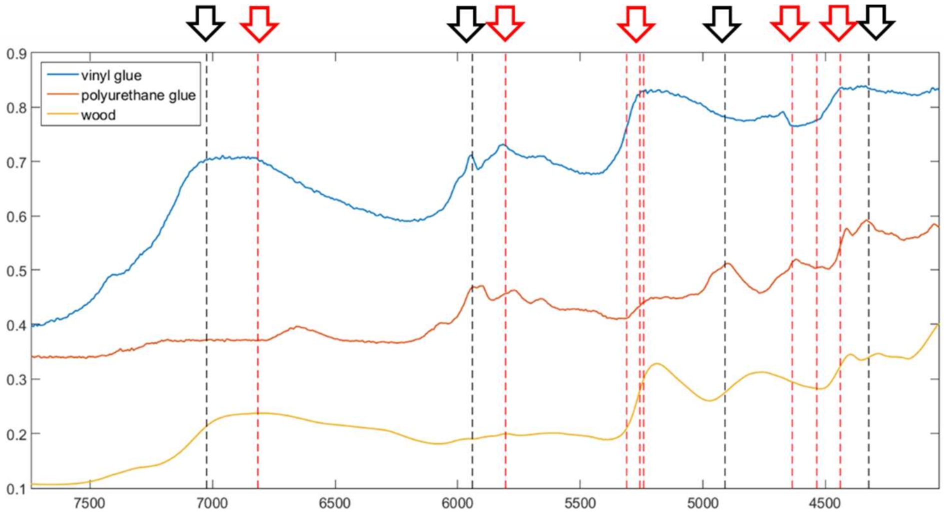

Figure 4-19: plot of the mean spectrum of virgin wood samples and of vinyl and polyurethane glue spectra. The black (wood) and red (glue) dotted lines show the most relevant wavenumbers.

Since the separation among the different groups is rather clear in the PCA score plot, a PLS-DA (Wold, Sjöström et al. 2001, Barker and Rayens 2003) model was applied for classification purposes. The data were pre-processed with the same pre-treatment already applied for PCA computation. The PLS-DA calibration model (n = 74 samples) was cross-validated using venetian blind-cross validation (5 segments) and an external test set (n = 32 samples) was used to validate the model. The training and test sets were chosen using the Kennard-Stone algorithm (Kennard and Stone 1969) and the metric used to compute the distance between points was the Euclidean. The classification model was developed using 5 LVs and present a high correlation ($Q^2 = 1.00$) and a low total misclassification rate of 10.4%. Furthermore, the values of 95% of sensitivity and 80% of specificity for training set and 95% of sensitivity and 86% of specificity for test set demonstrate the good classification accuracy of the model. The confusion matrix and a summary of the quality statistical parameters of the PLS DA model are given in Table 4-2. It should be noted that the confusion matrix reports the average samples (not the replicates), hence 'Class 0/1' indicates the number of samples, whose replicates were partly classified in class 1 and in class 0.

Table 4-2: confusion matrix and summary of the quality statistical parameters of the PLS DA model.

	Cross-Validation			**Prediction**		
	Class 1	Class 0/1	Class 0	Class 1	Class 0/1	Class 0
Class 1	42	5	0	17	3	0
Class 0	3	5	19	0	4	8
Misclassification rate	10.4%			8.3%		
Sensitivity	95%			95%		
Specificity	80%			86%		
LVs	5			5		

As showed in the PLS-DA score plot (Figure 4-20), the separation between treated and virgin wood is not so evident, as the separation between softwood and hardwood samples, located in the positive and negative values of LV2, respectively. The low percentage of glue in the glue-laminated wood brings to set these samples really close to the virgin wood, reducing the separation between the two groups. In order to increase the performance of the model, it could be worthwhile to consider virgin wood and glue laminated wood samples in a model

(MOD 2) and virgin wood and all the other treated wood samples in another one (MOD 1).

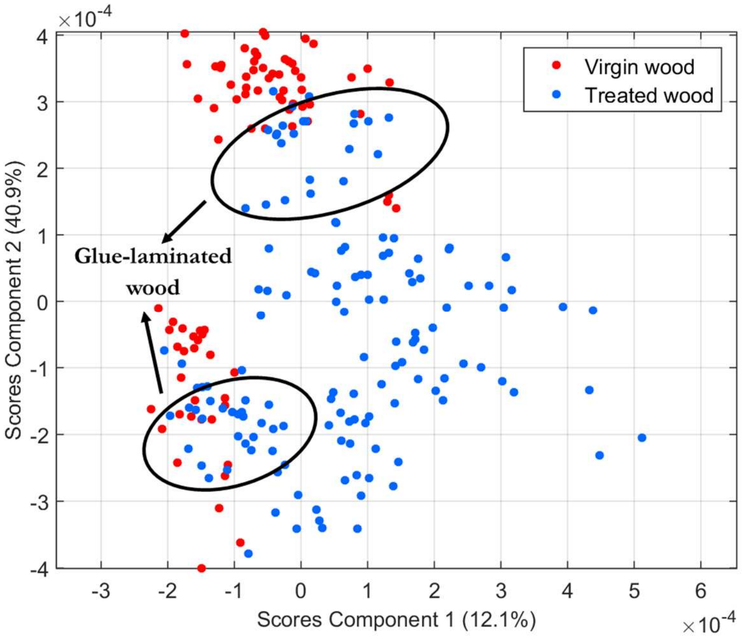

Figure 4-20: PLS-DA score plot of the two first LVs colored in accordance to treated and virgin wood. The two circles mark the area where glue-laminated samples are located.

MOD 1 was built with 56 samples (training set) and was tested using 24 samples (test set), while MOD 2 accounts 46 samples in the training set and 19 samples in the test set. The training and test sets were chosen using the Kennard-Stone algorithm (Kennard and Stone 1969).

As clear in Figure 4-21.a, MOD 1 is able to discriminate between treated and virgin wood samples and between hardwoods and softwoods, in particular the former have negative LV2 values and the latter positive LV2 values. Also the PLS-DA score plot of MOD 2 shows a good separation between hardwood and softwood samples: positive values of LV1 are related to softwood samples and LV1 negative values to hardwood samples (Figure 4-21.b). Instead, the treated and virgin wood samples are not so well separated, anyway a separation trend between the two classes is present and it is much more evident for softwoods than hardwoods. The greater dispersion of hardwood scores is probably due to the higher variability of hardwood in comparison to softwood, as already highlighted by the authors in **Paper** I (Chapter 4.1).

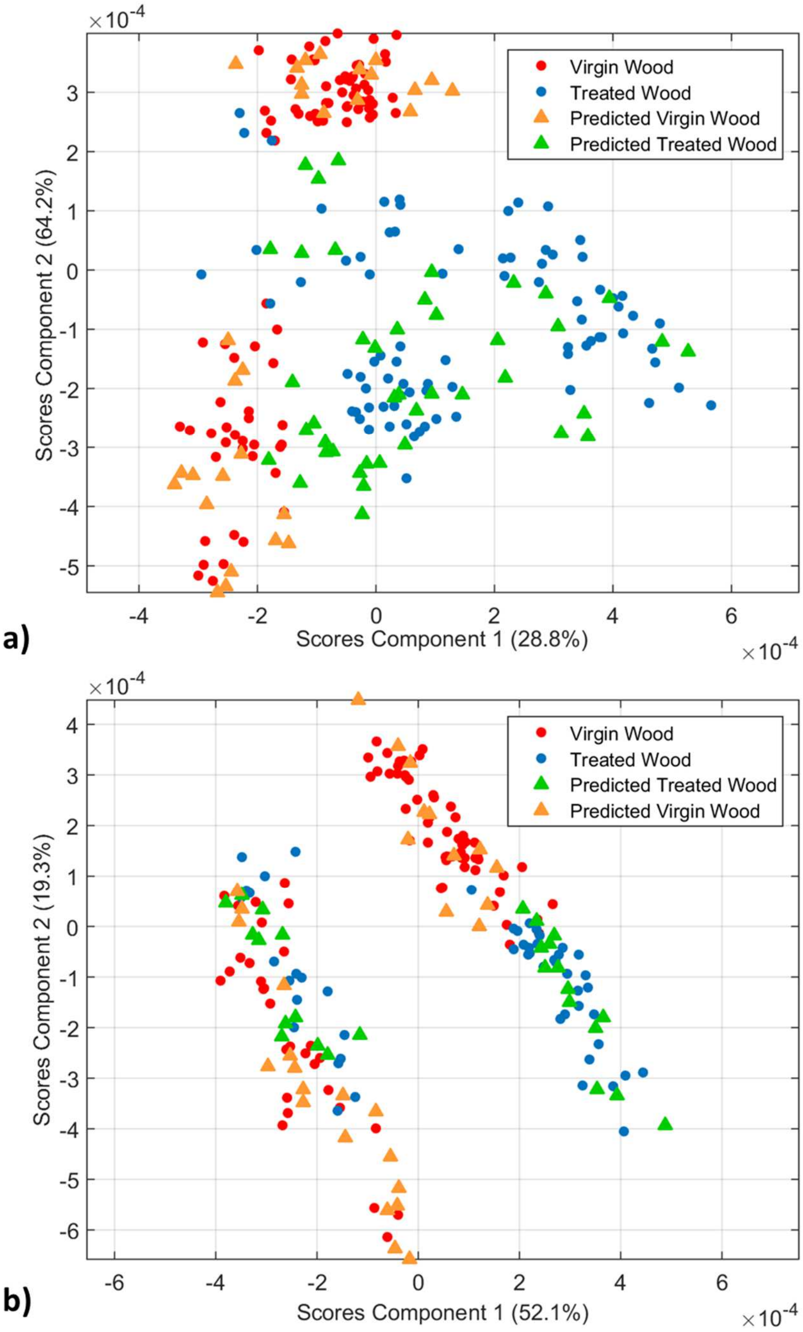

Figure 4-21: PLS-DA score plot of the model MOD 1 (a) and of the model MOD 2 (b). (●) Treated and untreated calibration samples. (▲)Treated and untreated validation samples.

Table 4-3 reports the confusion matrices and the summary of the quality statistical parameters of the two PLS DA models, in particular the number of latent variables used to develop the classification models, the sensitivity and specificity values and the misclassification rate. These values demonstrate that NIR spectroscopy has the potential to be a useful tool for the identification of treated wood. MOD 1 has high rate of sensitivity and specificity values, in particular it is able to correctly classify 100% for training set and 93% of the samples for the test set (sensitivity), and does not give many incorrect classifications for 100% for training set and 97% for test set samples (specificity). In addition, the few number of misclassification points out that the developed model could be used for classification purposes. MOD 2 presents worst classification performance than MOD 1. The low value of sensitivity in MOD 2 is probably due to the low percentage of glue present in glue-laminated wood samples and to the fact that the model separates mainly between hardwood and softwood samples rather than virgin and treated wood. Even so, the model can be used for screening classifications.

Trying to increase the accuracy of the model MOD 2, the dataset was split again on the basis of hardwood and softwood samples. A total of 38 and 26 samples were used for the development of the PLS-DA models for softwood and hardwood, respectively. Considering the reduced number of samples, both models were validated using venetian blind-cross validation (5 segments).

As shown in the two PLS-DA score plots (Figure 4-22), the separation trend is now better highlighted and related to treated/untreated samples. The classification models were developed using 4 LVs and have given back good results. The misclassification rate for softwood model is 4.4% and for hardwood model 0%. The sensitivity and specificity values are 93% and 97% for softwood model and 100% and 100% for hardwood model, respectively. The better classification accuracy of hardwood than softwood model could be due to the fact that, by removing hardwood and softwood information, the separation between treated / virgin wood is much more evident in hardwood than in softwood.

Table 4-3: confusion matrices and summary of the quality statistical parameters of PLS DA models (MOD 1 and MOD 2).

	MOD1						MOD2					
	Cross-Validation			Prediction			Cross-Validation			Prediction		
	Class 1	Class 0/1	Class 0	Class 1	Class 0/1	Class 0	Class 1	Class 0/1	Class 0	Class 1	Class 0/1	Class 0
Class 1	29	0	0	13	0	1	9	6	2	7	1	1
Class 0	0	0	27	0	1	9	0	1	28	1	2	7
Misclassification rate	0.0%			5.6%			11.6%			17.9%		
Sensitivity	100%			93%			73%			98%		
Specificity	100%			97%			98%			80%		
LVs	4			4			5			5		

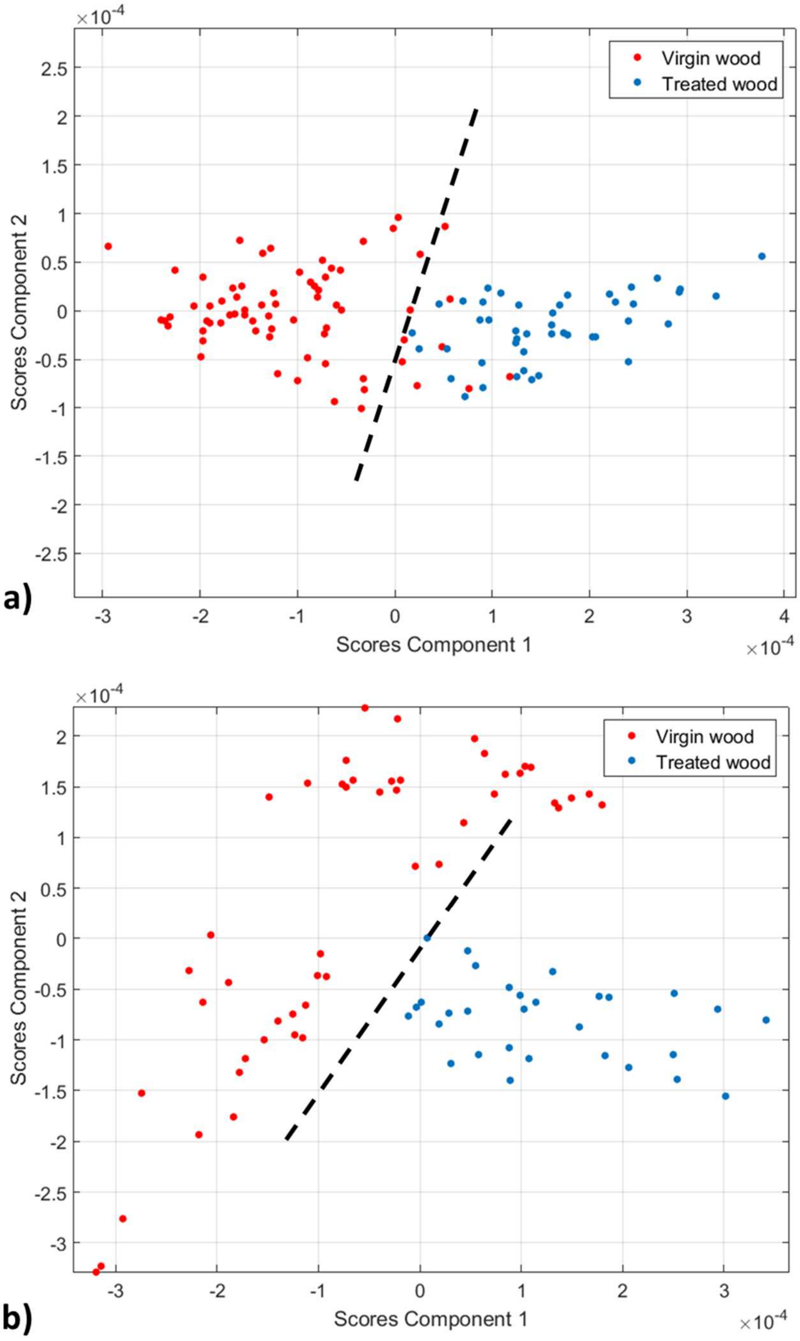

Figure 4-22: PLS-DA score plot of virgin and glue-laminated softwood samples (a). PLS-DA score plot of virgin and glue-laminated hardwood samples (b).

Different variables selection methods – regression vector, SR, VIP, iPLS and rPLS – were tested in order to simplify the models by reducing the number of variables and/or have better classifications. Considering MOD 1, the same classification performance of the total model was obtained using VIP, rPLS and forward iPLS method with number of intervals 50. In detail, rPLS and iPLS need less components for the development of the model (3 LVs instead of 4 LVs), while VIP requires more model parameters (5 LVs instead of 4LVs). The classification parameters of the fore mentioned models were calculated also for the test set. In this case, VIP method allows to get better results in classification – 94% sensitivity, 97% specificity and 5% misclassification rate – than rPLS and forward iPLS (sensitivity = 91%, specificity = 80% and misclassification rate = 14%). This indicates that iPLS and rPLS models simplify the model by reducing the number of wavelengths and of LVs, but they did not reduce the uncertainty.

A slight improvement of the classification ability of model MOD 2 can be observed using rPLS method. Indeed, the sensitivity value increased from 73% to 90%, the specificity value increased from 98% to 100% and the misclassification rate decreased from 11.6% to 3.6%. Moreover, a less number of LVs was used for modelling (from 5 LVs to 3 LVs). The rPLS classification parameters were calculated also for the test set and good results were achieved; sensitivity = 89%, specificity = 90% and misclassification rate = 11%.

The results presented suggest a profitable use of near infrared analysis coupled with multivariate data analysis for the classification of virgin wood and different types of residues from wood processing industry. The variables selection methods applied have returned an improvement of the classification performance of the two models when compared to the full-spectrum PLS-DA model.

The study demonstrated its potential especially if a ‘classification tree’ approach was followed. Indeed, the models shown in this study were obtained by splitting the dataset in accordance to the results of the previous partition. This procedure allowed to obtain models with a lower samples variability, but simpler and stronger in their classification. The results are still preliminary and the method might be improved increasing the number of samples, especially for the virgin and glue-laminated softwood/hardwood models. In any case NIR technique runs for being the first methodology able to provide information about the origin and source of the pellet, in particular about the presence of treated material in virgin wood pellet.

5. Infrared spectroscopy for quantitative analysis of biomass

---

The multiplicity of biofuels and their inherent and wide variability of the characteristics highlight the need to characterize their chemical and physical parameters. This knowledge is instrumental for finding the best valorisation of the product and for the development of a related market. The possibility to perform a rapid analysis of such characteristics is therefore fundamental and the results can support the correct energy exploitation of the material. Moreover, the industrial use of biofuels asks for rapid characterization methods appropriate for on-line applications.

As consequence, the overall aim of this chapter is to evaluate the potential of NIR spectroscopy technique to be applied as a tool to improve the quality control and have information about the main chemical and physical characteristics of the biomass.

5.1 Prediction of gross calorific value and ash content of woodchip samples *(Paper IV)*

Among the chemical-physical parameters set by the international standards EN ISO 17225, gross calorific value and ash content are two important key-elements for the determination of the biomass quality. In particular, ash content influences the combustion efficiency and could cause technological and environmental problems. Slagging in furnaces, fouling of heat exchanger surfaces (Kortelainen, Jokiniemi et al. 2015), bed agglomeration, corrosion in the combustion device (Werkelin, Skrifvars et al. 2010) and fly ash emissions (Narodoslawsky and Obernberger 1996) are only few examples. Moreover, it is really important for the discrimination among the different quality classes. Calorific value is an essential parameter in the specification of biomass quality to set the price of the product and is influenced by its chemical composition. It is also fundamental for the planning and control of power plants using biomass fuels (Posom, Shrestha et al. 2016). This chapter is based on the paper "Prediction of gross calorific value and ash content of woodchip samples by means of FT-NIR spectroscopy" accepted for publication in the peer-reviewed journal *Fuel Processing Technology*.

Several works have been already performed with the aim to develop models for the prediction of different parameters on specific kind of biomass or wood species, but the real strength point of this study was the samples collection approach. In fact, 125 woodchip samples were taken directly from the market. In detail, samples were collected from several power plants during long-term monitoring plan and from different parts of Italy, making them representative of the national scene and allowing the development of more robust and reliable prediction models.

All the samples were first stabilized at 45 °C for 24 h, ground down to 1 mm of particle size by means of a cutting mill and finally both the conventional and near infrared analyses were performed. GCV and Ac were determined in accordance to the analytical methodologies reported in the standards EN ISO 14918:2010 and EN ISO 18122:2015, respectively. The two parameters were calculated both air dried (GCV_{ad} and Ac_{ad}) and on a dry basis (GCV_{db} and Ac_{db}) by multiplying the air dried parameter by the dry matter content in percentage.

Different regression models were developed applying several pre-treatments including SNV, first and second derivative (Savitzky-Golay algorithm (Savitzky and Golay 1964) with 13 or 21 smoothing points and 2nd order polynomial). The models were validated using venetian blind-cross validation (5 segments). Considering the higher number of samples, PLS model for the prediction of Ac was also validated using an external test set. With the aim to maintain the representativeness of the training set, the test set was chosen *ad-hoc*. It was not possible to do the same for GCV prediction model because of the reduced number of samples.

Regarding Ac, good results were obtained considering both the parameter air dried and on a dry basis. Because of the high correlation between Ac_{ad} and Ac_{db} ($r= 0.999$) and the absence of differences observed between the prediction results of the two models, further elaborations were carried out only on Ac$_{ad}$, which is logical, as this is how the NIR instrument has seen the sample. The strongest model for the prediction of Ac$_{ad}$ was developed using no spectral pre-treatment, but applying only mean-centering. As no replicate outlier was found, each sample was averaged across the replicates. The reference values were also averaged. The 5 LVs PLS model has $R^2 = 0.81$, $Q^2 = 0.77$, RMSE = 0.42%, RMSECV = 0.47%, RPD = 2.1 and RER = 8.0, indicating the potential use of the model for screening applications (Figure 5-1 a).

Due to the low correlation between GCV$_{ad}$ and GCV$_{db}$ (r= 0.631), the prediction results of GCV$_{ad}$ were better than those on a dry basis and further elaborations have taken into account only the GCV$_{ad}$. The best model for the prediction of GCV$_{ad}$ was developed using first derivative (Savitzky-Golay with second-order polynomial and 21-points window) and calculating the average of the two replicates for each sample. The reference values were also averaged. A total of 5 LVs have been considered for building the model and the $R^2 = 0.89$, $Q^2 = 0.81$, RMSE = 217 J/g, RMSECV = 289 J/g, RPD = 2.5 and RER = 13.3 (Figure 5-1 b) indicate a good model that would be helpful as a screening tool for quality control application.

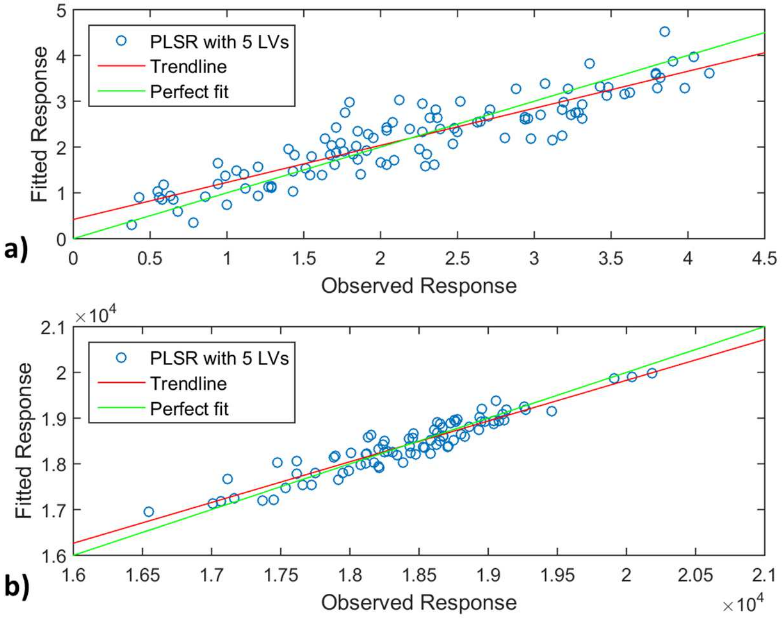

Figure 5-1: regression plots of observed versus fitted response of (a) ash content air dried and (b) gross calorific value air dried.

With the aim to enhance the performance of the prediction models in comparison to that of the full-spectrum models, different variables selection methods were applied, in detail: regression vector (B), Variables Important in Projections (VIP) (Wold, Sjöström et al. 2001, Chong and Jun 2005), Selectivity Ratio (SR) (Rajalahti, Arneberg et al. 2009), interval Partial Least-Squares (iPLS) (Norgaard, Saudland et al. 2000) and recursive weighted Partial Least Squares (rPLS) (Rinnan, Andersson et al. 2014). The results of the two best variables selection methods were reported in Table 5-1 and they referred both on the full FT-NIR spectral range (10,000-4,000 cm^{-1}) and on a selected wavelength range (7,500-4,000 cm^{-1}) containing the most significant information.

rPLS results to be the best variables selection method and, especially for gross calorific value, there is a significant improvement on the prediction performance. The GCV_{ad} model was developed using 7 LVs and considering the spectral range from 7,500 to 4,000 cm^{-1}. The RMSECV value decreased from 289 J/g to 234 J/g and Q^2 increased from 0.81 to 0.87. Regarding Ac_{ad} prediction, the rPLS model (5 LVs) was developed using the full spectrum since some structured information is contained also after 7,500 cm^{-1}. RMSECV value decreased from 0.47% to 0.44% and Q^2 increased from 0.77 to 0.80. As a consequence, both models could be used in screening applications for the biomass quality control. Considering root mean square error of prediction (RMSEP), rPLS model results to be similar to the PLS model and the B model on the spectral range from 7,500 to 4,000 cm^{-1}. This indicates that the rPLS does not reduce the uncertainty as such, but it reduces model complexity, thus making it easier to interpret.

Table 5-1: summary of variables selection results for Ac_{ad} *(%) and* GCV_{ad} *(J/g) prediction models. The best models are highlighted in bold.*

P*	**Model**		**Spectral range**	**Optimal LVs**	**Calibration**		**Cross validation**		**Prediction**	
	Type	n°I*			RMSE	R^2	RMSECV	Q^2	RMSEP	Q^2
Ac	PLS	-	10000-4000	5	0.42	0.81	0.47	0.77	0.49	0.78
	B	-	10000-4000	5	0.43	0.80	0.50	0.74	0.50	0.75
	rPLS (5LV)	**-**	**10000-4000**	**5**	**0.40**	**0.83**	**0.44**	**0.80**	**0.48**	**0.77**
	PLS	-	7500-4000	4	0.44	0.80	0.47	0.77	0.47	0.80
	B (5LV)	-	7500-4000	4	0.42	0.82	0.45	0.78	0.47	0.79
	rPLS (5LV)	-	7500-4000	5	0.40	0.83	0.44	0.80	0.49	0.77
GCV	PLS	-	10000-4000	5	217	0.89	289	0.81	-	-
	iPLS for (10LV)	50	10000-4000	6	227	0.88	258	0.85	-	-
	rPLS (7LV)	-	10000-4000	6	204	0.90	247	0.86	-	-
	PLS	-	7500-4000	5	221	0.89	286	0.82	-	-
	iPLS for (10LV)	29	7500-4000	5	207	0.90	237	0.87	-	-
	rPLS (7LV)	**-**	**7500-4000**	**6**	**195**	**0.91**	**235**	**0.87**	**-**	**-**

**** P = Parameter; n°I = number of interval.***

The variables selected by rPLS for GCV_{ad} and for Ac_{ad} models (Figure 5-2 b and Figure 5-3 b) and the regression vector plots of the total models (Figure 5-2 a and Figure 5-3 a) have been used to detect the most relevant variables, and associated compounds, responsible for the prediction of the two studied parameters.

The prediction of GCV is mainly linked to C-H bonds containing more energy than C-C bonds and O-H bonds (Lestander and Rhen 2005). Since it is well known that GCV and moisture content are negatively correlated, O-H bonds of H_2O are also important (Gillespie, Everard et al. 2015). Confirming the literature results, the wavelengths selected by rPLS (Figure 5-2 b) were mainly linked to C-H bonds (e.g. wavelength range 4497-4489 cm^{-1} and 4921-4902 cm^{-1}) (Schwanninger, Rodrigues et al. 2011, Popescu and Popescu 2013), a part from the range 5134-5168 cm^{-1} related to O-H stretching and O-H deformation vibration of water (Schwanninger, Rodrigues et al. 2011) and bands 7101-7104 cm^{-1} related to the 1st overtone of O-H stretching (Schwanninger, Rodrigues et al. 2011). At the end, it should be noted that not all the wavelengths already detected in the regression vector plot (Figure 5-2 a) seem to be significant for the GCV prediction. This means that rPLS not only performs clearly better than the global PLS model, but it is also useful in simplifying it, thus obtaining a model that is easier to understand/interpret.

Regarding Ac, the detection of the most relevant wavelengths – and related compounds – responsible for the prediction is more complex than for GCV model. Being the non-combustible portion of a biomass product (Bruun, Jensen et al. 2010), ash is difficult to predict since the inorganic constituent does not absorb in the near infrared region. Even if, inorganic compound can produce an infrared spectrum, the infrared bands for inorganic materials are broader, fewer in number and appear at lower wavenumbers than those observed for organic materials (Stuart 2005). According to Lestander and Rhen (Lestander and Rhen 2005), much of the inference relating to the prediction of Ac occurs in the carbon bonds. In any case, the prediction of Ac is an indirect prediction since it is based primarily on the presence of C-H and O-H bonds containing compounds (Gillespie, Everard et al. 2015).

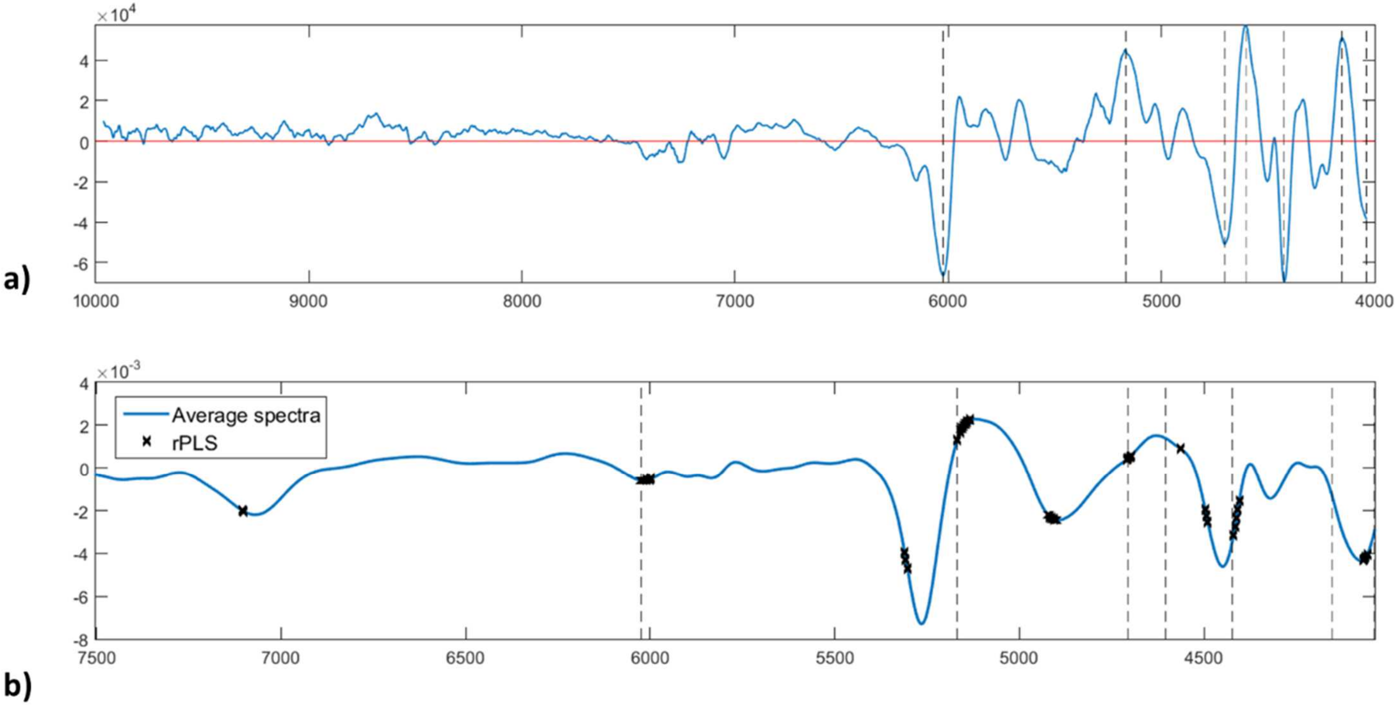

Figure 5-2: (a) PLS regression coefficient plot for GCV_{ad} *prediction with important wavenumbers indicated by dotted vertical lines. (b) Variables selected by rPLS for* GCV_{ad} *prediction. The dotted vertical lines indicate the wavenumbers detected as important in the PLS regression coefficient plot.*

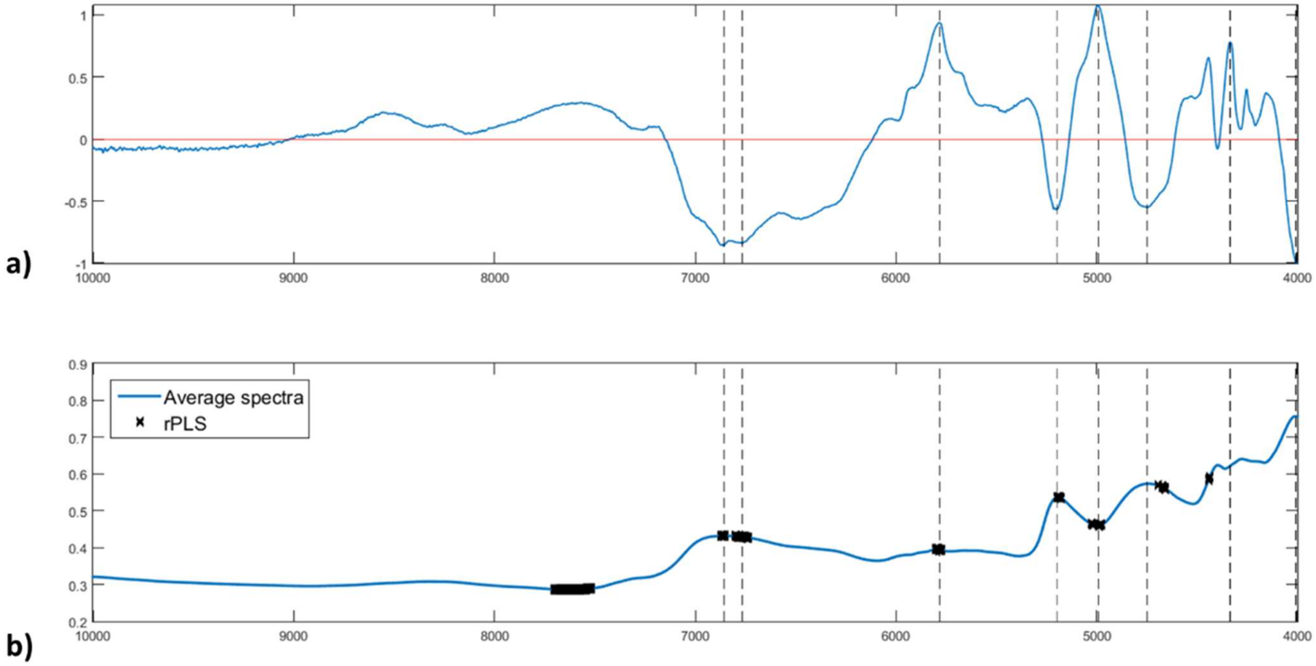

Figure 5-3: (a) PLS regression coefficient plot for Ac_{ad} prediction with important wavenumbers indicated by dotted vertical lines. (b) Variables selected by rPLS for Ac_{ad} prediction. The dotted vertical lines indicate the wavenumbers detected as important in the PLS regression coefficient plot.

Looking at Figure 5-3 b, peaks at 6858, 5793 and 4987 cm^{-1} are related to O-H bonds, in particular peak at 6858 cm^{-1} is assigned to 1st OT O-H stretching, 5793 cm^{-1} to 1st OT C-H of lignin (Schwanninger, Rodrigues et al. 2011) and 4987 cm^{-1} to O-H combination. The range 5188-5191 cm^{-1} is characteristic for the combination band of O-H stretching and O-H deformation vibration of water (Schwanninger, Rodrigues et al. 2011). Differently from the regression vector (Figure 5-3 a), rPLS model has selected some new variables in the first part of the spectra, in particular in the region between 7617 and 7563 cm^{-1} corresponding to 1st OT of C-H combinations.

The results of this study have demonstrated that FT-NIR technique coupled with PLS is a suitable tool for the prediction of GCV and Ac of woodchip samples and the models are exploitable in screening applications. In any case, it should be considered that the great utility of NIR method is based on the fast execution, reduced costs and simplicity of the analysis, but it does not replace the laboratory methods. In fact, such a tool could be used in evaluation-screening processes providing an indication of the biofuels quality and allowing a more accurate analysis of the biomass, not complying with the qualitative standards by using conventional analysis.

Moreover, the results showed the potential to develop a robust FT-NIR predictive model to be applied for industrial on-line applications. In this way, all the incoming feedstock in a power plant could be checked, providing a real-time quality trend of the product and overcoming sampling problems.

5.2 Influence of the scattering effects for the prediction of ash content (*Paper VII*)

Scattering consists in a deviation of light from a straight trajectory in one or more paths with no absorption. It is well known that this phenomenon distorts the data and when there is no interest in the physical artefacts of the spectra, but in the chemical information inherent in the data (far more frequent case), pre-processing must be applied. Multiplicative Scatter Correction (MSC) (Martens, Jensen et al. 1983), and its extended form Extended Multiplicative Signal Correction (EMSC) (Martens and Stark 1991), are well known pre-processing techniques capable in separating the physical light scatter from the chemical light absorption.

Trying to figure out whether it is the scattering effects, the chemical features or a mixture that really holds the information for the ash prediction, a more in-depth analysis was carried out on the woodchip dataset of **Paper** IV. The aim of the study was to investigate if the MSC/EMSC correction factors – and as a consequence the scattering effect – play an important role in the prediction of the ash content; a fundamental parameter for the biofuel sector. This chapter is based on a manuscript under finalization (**Paper** VII).

Two spectroscopic datasets have been taken into account: woodchips for the energy sector and pellet samples for the feed sector. The feed dataset was used as a test to demonstrate if it is really scattering that is important for modelling ash and if similar results could be obtained working on different matrices. The two datasets are different in typology of data and terms of samples and variables. The woodchip dataset consists of 109 samples analysed by Antaris II (Thermo Fisher Scientific Inc, USA) in duplicate in the spectral range from 10,000 to 4,000 cm^{-1}. The feed samples were measured by an InfraXact (Foss A/S, Denmark) instrument in the reflectance mode. The dataset consists of 229 samples acquired in the spectral range from 1,100 up to 1,848 nm.

The reference data is the ash content (%), a fundamental feature for both sectors. For energy sector ash content is one of the most important parameter, since inorganic matter influences the combustion efficiency in the furnace and is responsible of ash-forming elements, with resulting technological and environmental problems during biomass processing (Vassilev, Vassileva et al. 2017). In the feed sector, ash or mineral content gives indications of the energy content of dairy feed (Weiss 2017) and on the quality of the product and its nutritional parameters.

In order to assess which is the useful information for ash content prediction, NIR data were pre-processed by using MSC. Both original and pre-treated near infrared spectra were modelled by a PLS model (Wold, Sjöström et al. 2001). A five-segmented venetian blind cross-validation was used to validate the datasets.

A third PLS model (hereafter shortened to CORR-PLS) was developed by using the correction factors of MSC/EMSC and the scores of PCA. The procedure is briefly described: 1) the near infrared data were pre-processed using MSC or EMSC. The correction factors includes 0^{th}-2^{nd} order reference correction and none, up to 4^{th} order wavelength correction. The former refers to the order of the regression step, where normally MSC uses linear regression (1^{st} order reference correction) and the latter is the polynomial degree of the wavelength correction of EMSC. 2) A 20 components PCA model was developed on the mean-centered pre-processed full dataset. In order to 'standardize' the PCA solution and with the aim to reduce the errors due to the rotational uncertainty of the PCA, the PCA model on the whole dataset was used to rotate the PCA sub-models for the different segmentations (five cross-validation segments). In detail, Procrustes rotation (Hurley and Cattell 1962) was used to rotate the scores and loadings towards the global model. It should be noted that the reference spectrum for the MSC/EMSC is based on each segment, and not on the whole dataset. 3) The correction factors (i.e. the slope and intercept of MSC/EMSC) and the first 20 PCA scores were combined obtaining a new data matrix where the rows are the samples and the columns (variables) are the merged correction factors and PCA scores. All variables were 2nd-norm normalized in order to equalize the influence of each of the factors/ scores on the PLS model. 4) A PLS model was developed based on these normalized variables and the mean-centered ash content. 5) The left-out samples are predicted using the model from 4). Please note that the reference spectrum for the MSC/EMSC correction is based on the samples that are part of the PLS model from 4). This is also the case for the mean-centering of the spectral data, as well as the PCA.

The CORR-PLS models developed on the woodchip dataset provided better results than those obtained with PLS models on the raw spectra and on MSC pre-treated spectra. In particular, using 1st order reference correction and no correction for the wavelength, the CORR-PLS gave a $Q^2 = 0.80$ and RMSECV = 0.44% and only 3 LVs were necessary in the development of the model. Instead, PLS on raw data required the estimation of more model parameters (6 LVs) and returned $Q^2= 0.79$ and RMSECV = 0.45%, PLS model on MSC data needs 5 LVs and gave a $Q^2= 0.78$ and RMSECV = 0.46%. Figure 5-4 shows the RMSECV of the best CORR-PLS model compared with PLS model on raw data and MSC-model. It is interesting to note that the PLS model on the raw data performs similar to the one on the MSC corrected data. This strongly indicates that it is not necessary to perform pre-processing in order to get a good PLS model.

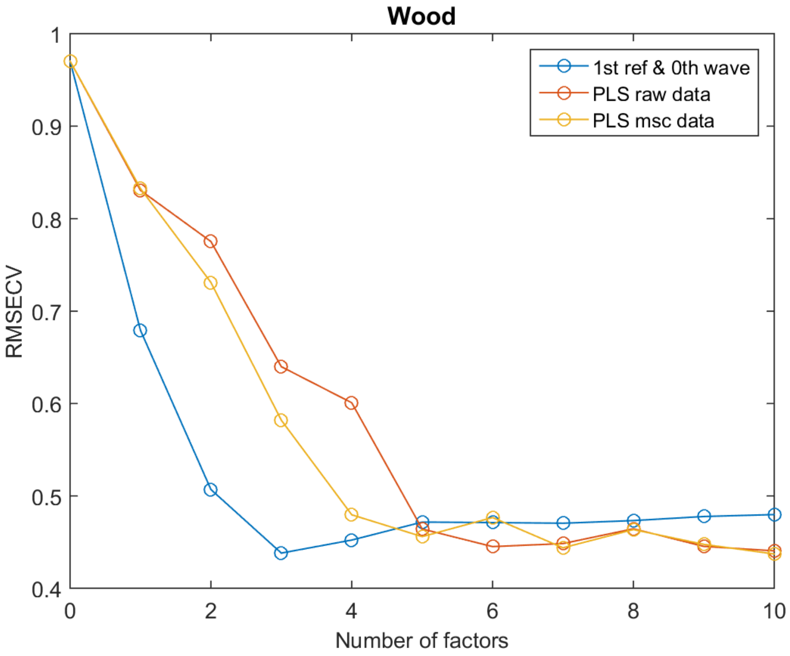

Figure 5-4: RMSECV of the PLS model on raw data, on MSC pre-treated data and CORR-PLS model for the woodchip dataset (ref: reference correction order, wave: correction order for wavelength).

Regarding the feed dataset, the prediction model on raw data was developed using 11 LVs and had Q^2 = 0.73 and RMSECV = 1.76%, while the MSC-model was developed using 10 LVs and returned Q^2 = 0.77 and RMSECV = 1.64%. Considering the lower number of LVs (3 LVs) and the good results (Q^2 = 0.78 and RMSECV = 1.61%), CORR-PLS model with 1st order reference correction and 1st order correction for the wavelength is considered the optimal prediction model, see Figure 5-5. The lowest RMSECV of the CORR-PLS model was the one using 1st order reference correction and 3rd order correction for the wavelength, with a Q^2 = 0.78 and RMSECV of 1.60% (4 LVs). However, due to the increased complexity of the model, and the small improvement, it was decided that the slightly worse CORR-PLS (with 1st order reference and 1st order correction for the wavelength) was the optimal.

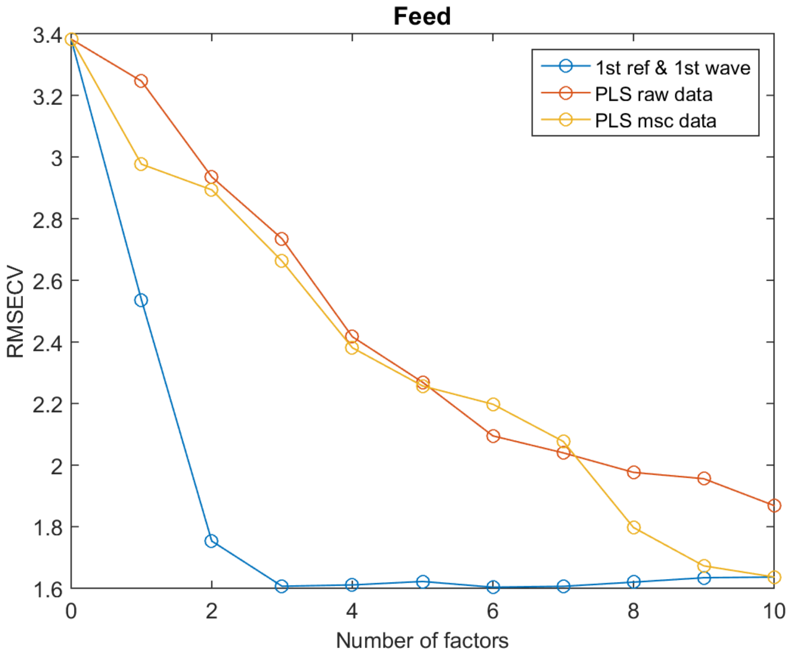

Figure 5-5: RMSECV of the PLS model on raw data, on MSC pre-treated data and CORR-PLS model for the feed dataset (ref: reference correction order, wave: correction order for wavelength).

As already mentioned, ash is not very active in the near infrared spectrum because of the low absorption of the molecules in this electromagnetic region and, by association, the prediction of this parameter using NIR spectroscopy is not easy. Furthermore, ash is not a well-defined attribute, and as such it is difficult to assess what wavelengths would be most appropriate for the PLS model.

The regression coefficient plots of the CORR-PLS models were investigated in order to figure out if the models are focusing on the same chemical information of the PLS models developed on the raw data or MSC pre-treated data, or rather are looking at different areas in the NIR spectra. The ultimate goal is to verify if scattering effects hold useful information for the ash content prediction.

Looking at Figure 5-6 it is interesting to note that the regression vectors of all the models – PLS on raw data, on MSC data and CORR-PLS – mainly look at the same variation in NIR. Going into detail, some small differences can be observed in the woodchip model. The CORR-PLS does not use the peak at 5199 cm^{-1}, but rather the peak at 6858 cm^{-1}. This latter is less influential in the PLS model on raw data and MSC data than in CORR-PLS model. In addition, another difference can be noticed at 4748 cm^{-1}, CORR-PLS model presents a shift and the peak falls at 4648 cm^{-1}. The regression vectors of the three feed models also look similar, except for differences mostly linked to the amplitude of the peaks rather than to the position.

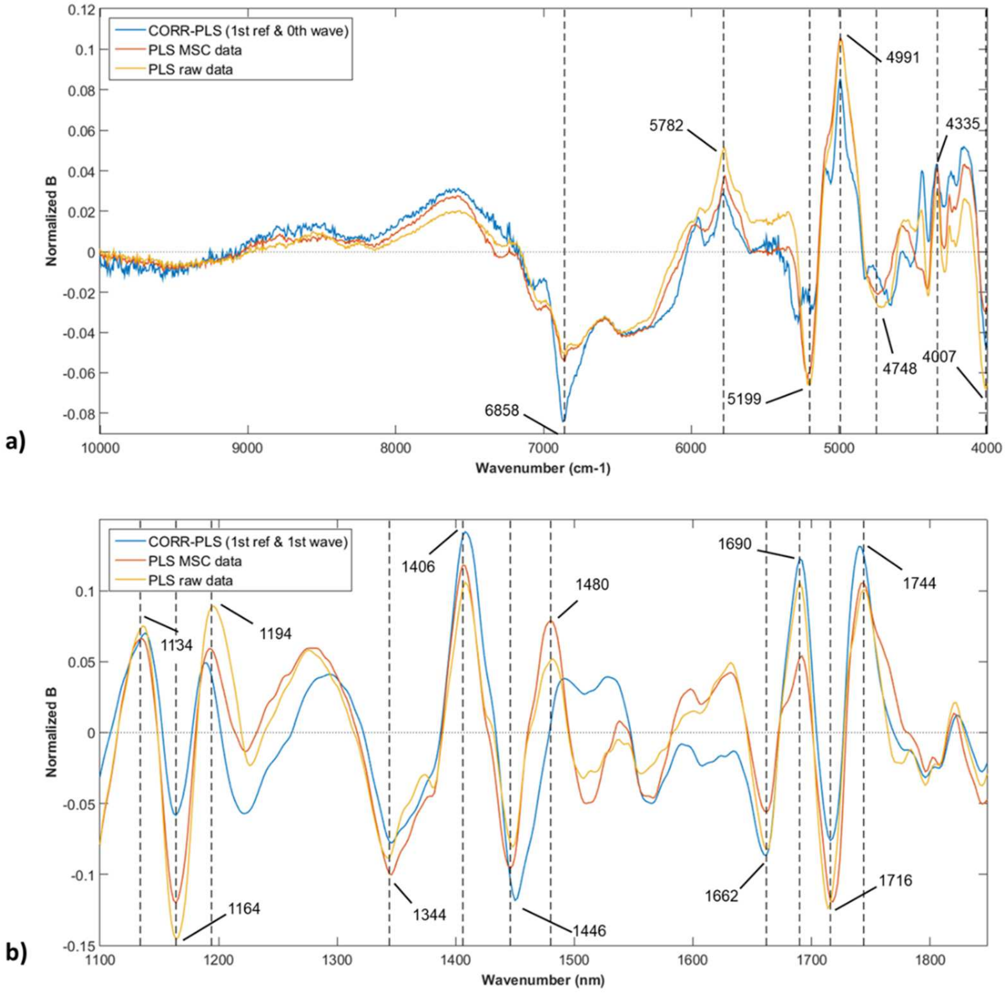

Figure 5-6: regression vectors of the CORR-PLS model (blue line), MSC-PLS model (red line) and PLS model on raw data (yellow line) of woodchip samples (a) and feed samples (b). The important wavenumbers for the prediction of ash content are marked in the plot with dotted lines.

Figure 5-7 shows the contribution of slope, intercept and PCA scores in the CORR-PLS model and confirms that the scattering properties are important, and furthermore indicates that it is not only the major variation in X that is essential in the prediction.

Summarizing, the use of the MSC/EMSC correction factors for the prediction of ash content turns out to be fundamental and allows to obtain a simpler model. Less components are needed in the CORR-PLS model when compared to the regular PLS models getting even a slight improvement of the performance. Hence, it is clear from the results that scattering properties are responsible for the prediction ability of this parameter TOGETHER with chemical information.

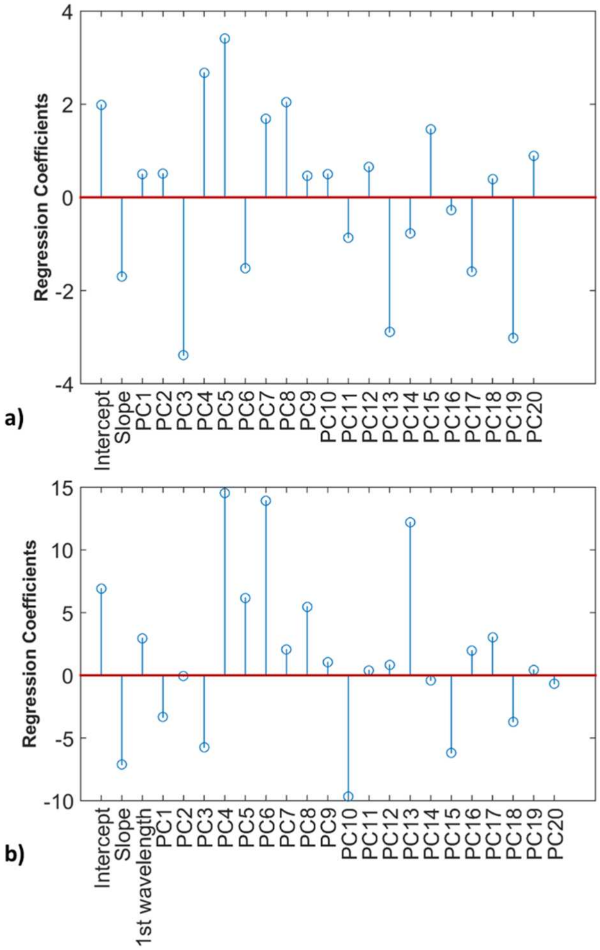

Figure 5-7: CORR-PLS regression vectors of woodchip (a) and feed (b) models showing the contribution of MSC/EMSC correction factors and first 20 PCA scores.

5.3 Laboratory customized on-line measurements for the prediction of the key-parameters of biomass

In order to perform a more effective monitoring of the biomass, NIR spectroscopy can be applied also on-line. The key parameters of biomass quality, such as moisture content, ash content and gross calorific value could be easily predicted directly in the production line.

In different fields, on-line NIR spectrophotometers are already used in the production line, providing continuous measurements and giving back the quality trend of the product and the possible critical points, in real time. In the biofuel sector, the great advantage of this technique is the implementation of the automatic process of control on a great mass of biomass, e.g. all the incoming feedstock in a power plan, allowing a total quality control. This approach avoids the sampling procedure and the related problems with the product representativeness. Moreover, on-line control could improve the process efficiency and help in saving money; in fact the material not-conforming to the quality standards of the power plant would be not accepted as biofuels.

Considering the good results achieved using different laboratory spectrophotometers, in this study an on-line instrument customized for laboratory analysis was employed to test the performance of NIR spectroscopy, simulating the production line. In particular, on-line NIR technique was applied in order to predict GCV and Ac of wood samples, both in ground and in pellet form. Woodchip samples were also analysed as received with the aim to predict Mc of the material without any process. Moreover three mini-tasks were carried out in order to verify the ability of the technique to detect for differences in the origin and source of the material, in particular i) herbaceous/woody biomass, ii) hardwood/softwood, iii) virgin/treated wood. The results of this Chapter are not be submitted as a paper at the present.

5.3.1 Materials and methods

Different samples have been collected in order to get information about the origin and source of the material, and in particular:

- mini-task herbaceous/woody biomass (MT1): 15 samples of herbaceous biomass (maize, sorghum and straw) and 5 samples of woodchip,
- mini-task hardwood/softwood (MT2): 10 samples of fir, 10 samples of pine and 8 samples of beech,
- mini-task virgin/treated wood (MT3): 28 virgin wood samples (the same used also for the mini-task hardwood/softwood) and 10 treated wood samples (4 chipboard samples and 6 samples between OSB and plywood).

For the prediction of the chemical-physical parameters, different types of material were also collected and analysed by NIR spectroscopy. In particular 58 woodchip samples as received were used for the development of the Mc prediction model and 38 wood pellet samples, as received and after grinding, for the prediction of GCV and Ac. The ground material simulates the dried particles immediately before pelletizing process.

A part from pellet and woodchip, directly analysed in their original form, all the other samples were ground to below 1 mm of particle size by means of a cutting mill (mod. SM 2000, RETSCH).

The analytical methodologies adopted for the determination of GCV and Ac refer to the standards EN ISO 14918:2010 and EN ISO 18122:2015, respectively. Both parameters were considered on a dry basis, since it is the result actually required by the standard. The moisture content of the material as received was performed by using the analytical methodology according to EN ISO 18134-2:2015. All the mentioned methodologies were already described in Chapter 2.2.

This task was carried out using an on-line NIR spectrophotometer loaned by Büchi (NIR-Online; Büchi Labortechnik AG, Flawil, Switzerland). The instrument has a tungsten halogen dual lamp, is based on diode-array technology and collects spectral data from material placed in a rotating module located below the spectrometer (X-Rot module; Büchi Labortechnik AG, Flawil, Switzerland), simulating a conveyor belt in a pellet or power plant. The instrument was set to acquire the spectra at a wavelength range from 400 to 1,700 nm (i.e. also part of the visible wavelength range was included), 32 scans per sample and a spectral resolution of 5 cm^{-1}. In order to exclude the signal not associated to differences among samples but to the environment, a blank spectrum was acquired at the beginning of each analysis session. Spectra were recorded at room temperature

(18-20°C) and in duplicate for each sample. The acquisition window is 4 cm diameter.

Chemometric techniques and multivariate data analysis were used in order to extract information from spectral data about the origin and source of the material (PCA) and to develop models for the prediction of Mc, GCV and Ac (PLS). As a first step, the raw spectra acquired were pre-treated and the "optimal" pre-treatment was chosen for each PLS or PCA models developed. In particular SNV and first and second derivative (Savitzky-Golay method (Savitzky and Golay 1964) with different segment size and second polynomial order) were tested. Successively, PLS models were developed for the prediction of GCV, Ac and Mc, while PCA was computed in order to search for spectral differences between the different types of biomass analysed (herbaceous/woody material, hardwood/softwood, treated/untreated wood).

5.3.2 PCA results

PCA computations were performed on the spectral dataset pre-treated using second derivative (Savitzky-Golay method (Savitzky and Golay 1964) with 13 smoothing points and second polynomial order). To take into consideration that the visible part of the electromagnetic spectrum (400-780 nm) was removed from the MT2 PCA computation since it does not contain relevant information for the discrimination between hardwood and softwood samples.

Figure 5-8 shows the PCA scores plots of MT1 (a), MT2 (b) and MT3 (c). As it can be noticed from Figure 5-8.a, there is a separation between herbaceous and woody biomass along PC2: woodchip samples have negative PC2 scores whereas herbaceous biomass has positive PC2 scores. Moreover, there is a separation between straw and maize – positive PC1 scores – and sorghum – negative PC1 scores. This is due to the fact that sorghum samples are easily detected by the visible part of the electromagnetic spectrum (400-780 nm) since the ground material presents red particles deriving from its culm.

Figure 5-8.b shows the separation between hardwood and softwood samples. In particular, there is a separation along PC2: softwood samples have positive PC2 scores, whereas hardwood samples have negative PC2 scores. It is also possible to observe a separation between fir and pine samples along PC1.

At the end, Figure 5-8.c reports the PCA scores plot of treated and virgin wood samples. A clear separation between chipboard and the rest of samples could be observed, while OSB and plywood samples are close to the virgin wood. This is probably related to the lower quantity of glue used for these types of wood-composite materials.

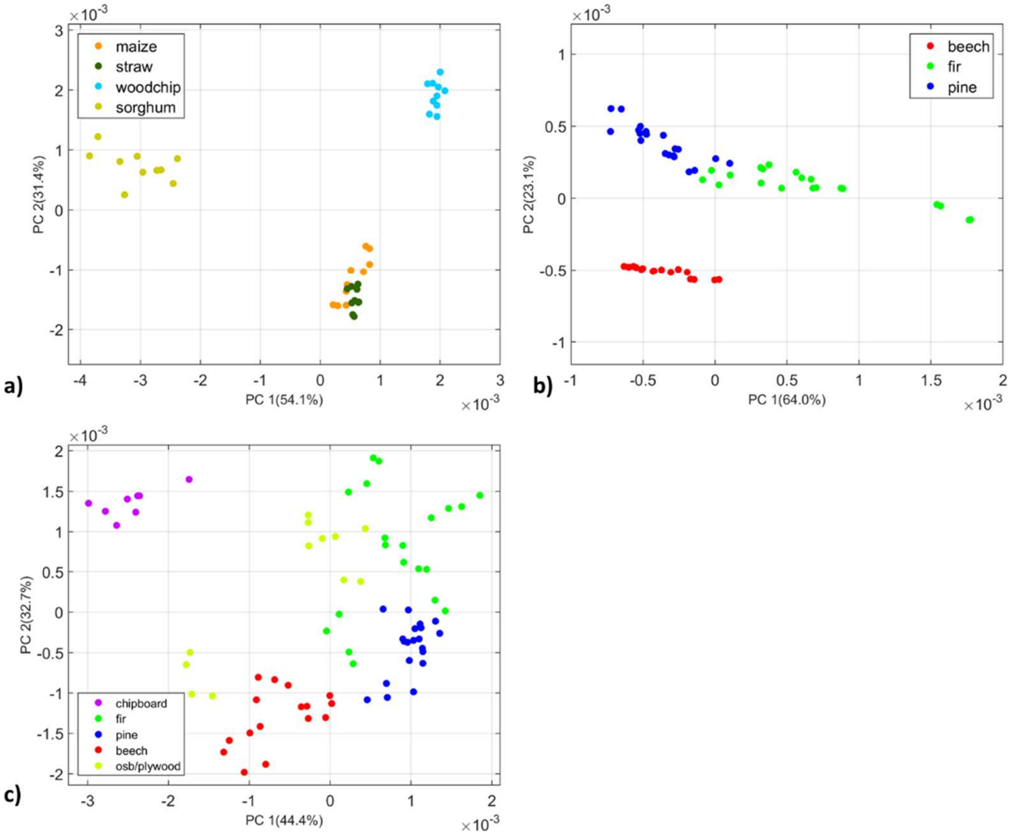

Figure 5-8: PCA scores plots of (a) herbaceous/woody biomass (MT1), (b) hardwood/softwood (MT2), (c) treated/virgin wood (MT3).

5.3.3 PLS results

A general descriptive statistic of the three parameters of interest is reported in Table 5-2. The results were obtained from the computation of all the samples both for Mc (58 samples) and Ac and GCV (38 samples) parameters. The woodchip samples show a huge variability in moisture content, ranging from 2.5%

to 56.1%. Pellet samples present a smaller variability in Ac and GCV range and, by association, a lower standard deviation.

Table 5-2: general descriptive statistic of woodchip and pellet samples (ar: as arrived, db: dry basis).

	Mc (% $_{ar}$)	GCV (MJ/kg $_{db}$)	Ac (% $_{db}$)
Mean	31.3	20.1	1.25
Standard deviation	15.9	0.4	1.12
Min	2.5	19.4	0.34
Max	56.1	20.8	5.16
Range	53.6	1.4	4.82

Table 5-3 reports the PLS results of the different pre-treatments carried out on the spectral dataset of woodchip samples for the prediction of Mc. PLS models were all validated using venetian blind-cross validation (5 segments). The best prediction model was developed using 7 LVs and first derivative (21 smoothing points with Savitzky-Golay method (Savitzky and Golay 1964), 2nd polynomial order) as pre-treatment. The high value of Q^2 (0.99) and the RMSECV (1.49%), RPD (10.7) and RER (36.0) values indicate the good prediction accuracy and the model could be used in any quality control application.

Table 5-3: summary of the PLS regression models developed for the prediction of moisture content (21der1: first derivative with 21 number of smoothing points; 21der2: second derivative with 21 number of smoothing points).

Pre-treatment	LVs	RMSE (%)	R^2	RMSECV (%)	Q^2	RPD	RER
No pre-treatment	6	1.65	0.99	2.01	0.98	7.9	26.7
SNV	5	2.38	0.98	2.79	0.97	5.7	19.2
21der2	5	1.34	0.99	1.56	0.99	10.2	34.4
21der1	**7**	**1.31**	**0.99**	**1.49**	**0.99**	**10.7**	**36.0**

PLS models were developed also for the prediction of GCV and Ac of 38 wood samples, analysed ground or directly in the pellet form. Considering the limited number of samples, the models were validated using Leave-One-out Cross validation.

The results of Ac and GCV predictions are shown in Table 5-4 and

Table 5-5, respectively. In general, the best results for the Ac prediction were obtained analysing the material in its pellet form. The best prediction model was developed using SNV as pre-treatment and 6 LVs. Returning $Q^2 = 0.81$, RMSECV = 0.49% RPD = 2.3 and RER = 9.8, it indicates the feasibility to use the model for screening quality applications. The prediction of Ac directly in its pellet form has the great advantage to analyse the sample without any operations (e.g. stabilization or grinding) and, as a result, the technique could be applied directly in line after the pelletizing process.

In order to figure out if the visible part of the spectrum (400-780 nm) carried out important information for the Ac prediction, the same model was developed also removing this part of the spectrum. A total of 4 LVs were used for the development of the model, Q^2 value increased from 0.81 to 0.87, while RMSECV decreased from 0.49% to 0.39%, demonstrating that the wavelengths from 400 up to 780 nm contain non-relevant variation for the prediction of this parameter.

Table 5-4: summary of the PLS regression models developed for the prediction of ash content (21der1: first derivative with 21 number of smoothing points; 21der2: second derivative with 21 number of smoothing points).

			Pellet				
Pre-treatment	LVs	RMSE (%)	R^2	RMSECV (%)	Q^2	RPD	RER
No pre-treatment	6	0.43	0.85	0.56	0.75	2.0	8.6
SNV	**6**	**0.32**	**0.91**	**0.49**	**0.81**	**2.3**	**9.8**
21der1	7	0.31	0.92	0.51	0.79	2.2	9.5
21der2	1	0.69	0.61	0.77	0.52	1.5	6.3
			Ground material				
No pre-treatment	8	0.35	0.90	0.61	0.71	1.8	7.9
SNV	6	0.45	0.84	0.65	0.67	1.7	7.4
21der1	5	0.51	0.79	0.65	0.67	1.7	7.4
21der2	1	0.79	0.50	0.87	0.39	1.3	5.5

Considering GCV, two outliers were found and have been deleted from the dataset before PLS computation. The best prediction model was developed using second derivatives (21 smoothing points with Savitzky-Golay method (Savitzky and Golay 1964), 2nd polynomial order) as pre-treatment directly on the material in its pellet form. The model uses 5 LVs and the Q^2 (0.77), RMSECV (0.17 MJ/kg), RPD (2.4) and RER (8.2) values indicate the potential use of the model for screening applications.

As for Ac, also GCV predictive model was developed removing the visible part of the spectrum (400-780 nm) and using the same pre-treatment of the full-spectrum model. For the prediction model directly on the pellet samples, 4 LVs were used and Q^2 value increased from 0.77 to 0.83, while RMSECV decreased from 0.17 MJ/kg to 0.15 MJ/kg. Instead, for the prediction model on the ground material, 4 LVs were used and Q^2 value increased from 0.71 to 0.77, while RMSECV decreased from 0.19 MJ/kg to 0.17 MJ/kg. Also in this case, the visible part of the spectrum is irrelevant for the prediction of the parameter.

Table 5-5: summary of the PLS regression models developed for the prediction of gross calorific value (21der1: first derivative with 21 number of smoothing points; 21der2: second derivative with 21 number of smoothing points).

Pellet							
Pre-treatment	LVs	RMSE (MJ/kg)	R^2	RMSECV (MJ/kg)	Q^2	RPD	RER
No pre-treatment	6	0.14	0.84	0.20	0.71	2.0	7.0
SNV	5	0.19	0.73	0.24	0.56	1.7	5.8
21der1	4	0.14	0.84	0.18	0.74	2.2	7.8
21der2	**4**	**0.12**	**0.89**	**0.17**	**0.77**	**2.4**	**8.2**
Ground material							
No pre-treatment	4	0.20	0.69	0.24	0.58	1.7	5.8
SNV	3	0.22	0.62	0.25	0.5	1.6	5.6
21der1	3	0.19	0.72	0.23	0.61	1.7	6.1
21der2	2	0.15	0.81	0.19	0.71	2.1	7.4

The preliminary results obtained are really encouraging, suggesting that the development of a technique based on chemometrics and NIR spectroscopy for providing early evidence of non-conformity and for predicting the main quantitative parameters of the biomass would be possible.

6. Conclusions and perspectives

The work presented in this Ph.D. thesis aims at answering the requests of the technical standards EN ISO 17225. According to these standards, the biomass quality is determined not only by chemical-physical parameters, but also by qualitative attributes. Unfortunately, information on biofuels is difficult to be achieved by conventional analysis or microscopy techniques because they are expensive, time-consuming and require specialized operators. Especially the quality of densified products, such as pellet, is hard to be assessed since they are made up with ground feedstock, becoming impossible to identify the typology of the original raw materials by naked eye. A suitable alternative comes by infrared spectroscopy which is already widely used for quantitative and qualitative purposes in several fields.

Different examples of infrared spectroscopy applications in the field of biofuels characterization are presented in this Ph.D. thesis. In particular, the infrared technique was used to identify the origin and source of the material. Going into detail, a SIMCA discrimination model was developed using FT-IR spectroscopy to provide first indications about the hardwood/softwood composition of the biomass. The results are really encouraging and the model is able in recognizing 93% of hardwood samples, 100% of softwood samples and 100% of the tested blends (**Paper** I). In addition, FT-IR analysis coupled with PLS regression turned out to be a very efficient tool in predicting hardwood and softwood contents in blend samples with a standard error of few points (3.8%) (**Paper** II).

A NIR model for the discrimination between hardwood and softwood and between bark and wood is already efficient and able in discerning three of the most important species in the pellet sector (fir, pine and beech) (**Paper** III). Such a tool could be fundamental for the pellet quality control, since the model developed is able to recognize 20% bark content in blend samples at a 99% confidence interval. Moreover, new perspectives could be imagined. The linear trend in the blends along the axis of pure wood to pure bark suggests that building a regression model with the purpose of estimating the amount of bark in wood samples is fairly straightforward. In addition, the good separation between the different classes in the PCA score plot suggests that the development of a discrimination model (PLS-DA or SIMCA) is also possible.

A further step in retrieving information about the origin and source of the material was obtained building a classification model for the detection of wood-based composite materials (i.e. glue-laminated wood, plywood, oriented strand board and chipboard) in the pellet sector using NIR spectroscopy. Considering the first exploratory study results, two PLS-DA models were developed: the former considering glue-laminated/virgin wood and the latter all the other treated wood samples/virgin wood. In particular, the first model shows 100% correct classification and the second model only 3.6% misclassification rate confirming the profitable use of NIR spectroscopy in getting information about the origin and source of the biomass (**Paper** V and **Paper** VI).

Regarding quantitative information, regression models for the determination of gross calorific value and ash content of woodchip samples have been developed and variables selection techniques have been used to improve the prediction performance. Comparing to the different variables selection methods applied, rPLS results to be the best one. In particular GCV was predicted with a RMSECV of 234 J/g ($Q^2 = 0.87$) and Ac with a RMSECV of 0.44% ($Q^2 = 0.80$), making the model exploitable in screening applications. The results demonstrated that FT-NIR is a suitable tool both for the prediction of GCV and Ac of woodchip samples and showed the potential to develop a robust predictive model for industrial on-line applications (**Paper** IV). In-depth analysis was carried out for the prediction of ash content and the results have demonstrated that scattering, together with chemical information, effects the prediction of such parameter. Working with a PLS model based on the MSC/EMSC correction factors and PCA scores of the pre-treated dataset, it is possible to obtain a slight improvement of the regression model and the decrease complexity of the system (**Paper** VII).

Furthermore, a preliminary study was carried out to investigate the possibility to apply NIR spectroscopy directly in the production line for the biomass quality control. Models for the prediction of the main quantitative parameters of the biomass (ash content, calorific value and moisture content) and for the classification of the type of material (wood/herbaceous biomass, hardwood/softwood, virgin/treated wood) were developed using an on-line NIR spectrophotometer. The first results are encouraging and give contributions to the prediction of the main chemical-physical parameters and the identification of the origin and source of the biomass, suggesting the feasibility to perform a quality control directly in the production line.

The outcomes of this Ph.D. thesis are really promising and highlight different possible benefits for the sector. Compared with conventional chemical analysis, infrared spectroscopy allows to check simultaneously with a single and fast measurement the quality parameters requested by EN ISO 17225. In addition, the control is not just about the chemical-physical parameters, but also the qualitative attributes, at the moment only declared by the producers and almost impossible to be monitored by traditional analysis. Since biomass presents an inherent variability and a structural complexity, a proper quality characterization is deeply linked to a correct sampling procedure, but it requires specialized operators as for the fore mentioned conventional analysis. Of course, experienced person working in the lab translates into additional costs for the industry and it does not fit with the economic return of the sector.

The strong reduction in cost and time of infrared analysis promotes the spread of quality control. This leads in a more effective and extensive monitoring of the bioenergy chain; a higher number of material could be analysed solving the problems of manual sampling and its related representativeness, fundamental aspects in highly heterogeneous matrices. It follows in the feasibility to trace and verify more efficiently the origin (e.g. wood species) and the type of biomass (e.g. bark / wood). But, even more important, it enables the control of the presence of treated wood (e.g. residues from wood processing industry) especially in densified products, such as pellets and briquettes. This is particular relevant since in different European countries, and especially in Italy, these materials are considered as waste and the energy use is normally only allowed in big plants, authorized and dedicated for waste treatment.

The application of the technique in hand-held portable devices or directly in the production line is the real strength. A huge amount of information useful for biofuels assessment could be collected with several advantages for the sector.

Indeed, the more effective quality control of the biofuel used has positive effects for environmental sustainability assessments and translates into the potential reduction in combustion emissions. Excluding non-conforming and low quality products (e.g. wood residues with glue, high ash content) is fundamental to ensure lower environmental impact levels. In fact, information about biomass quality allows to tune the combustion process in function of the fuel supplied. Moreover, the technique promotes for sure the total quality control of the product. This translates in the possibility to assess large bulk of biomass (e.g. plant biomass pile, biomass shipload, feedstock supply in pelletizing/power plant) without having problems of manual sampling.

It is therefore clear that the development of a tool based on IR spectroscopy will open up new opportunities for the biofuel sector, supporting its monitoring at different levels of the supply chain (production, distribution and end-use) and providing an indication of the product quality useful in evaluation-screening process. To be taken into consideration that infrared spectroscopy will not completely replace conventional analysis, but will allow a more extensive biomass control identifying the critical points in the bioenergy chain and will help to achieve cost-savings. As consequence, all the stakeholders - solid biofuel producers, traders but also customers - can have benefits in the use of this new system of quality control.

Further steps in the development of a procedure based on chemometrics and infrared spectroscopy for the quality control could be done. In particular, the prediction performance of the most important chemical-physical parameters characterizing the biofuel quality class could be improved. Moreover, additional studies should be carried out with the purpose to discriminate the qualitative attributes of different types of biomass, e.g. woody/herbaceous biomass, determination of bark content in wood samples. As last, the results of the Ph.D. activities suggest that the development of a sector-specific technical standard that works on infrared spectroscopy is fairly straightforward and the future activities plan will go for sure in this direction.

7. References

Afseth, N. K. and A. Kohler (2012). "Extended multiplicative signal correction in vibrational spectroscopy, a tutorial." *Chemometrics and Intelligent Laboratory Systems* **117**: 92-99.

Andersen, C. M. and R. Bro (2010). "Variable selection in regression—a tutorial." *Journal of Chemometrics* **24**(11-12): 728-737.

Bajpai, P. (2016). *Structure of Lignocellulosic Biomass*, in "Pretreatment of Lignocellulosic Biomass for Biofuel Production". Singapore, Springer Singapore. 7-12.

Ballabio, D. and R. Todeschini (2009). *Multivariate Classification for Qualitative Analysis*, in "Infrared Spectroscopy for Food Quality Analysis and Control", Elsevier Inc.

Barker, M. and W. Rayens (2003). "Partial least squares for discrimination." *Journal of Chemometrics* **17**(3): 166-173.

Bjarnestad, S. and O. Dahlman (2002). "Chemical Compositions of Hardwood and Softwood Pulps Employing Photoacoustic Fourier Transform Infrared Spectroscopy in Combination with Partial Least-Squares Analysis." *Analytical Chemistry* **74**(22): 5851-5858.

Blanco, M. and I. Villarroya (2002). "NIR spectroscopy: a rapid-response analytical tool." *TrAC Trends in Analytical Chemistry* **21**(4): 240-250.

Bro, R. and A. K. Smilde (2014). "Principal component analysis." *Analytical Methods* **6**(9): 2812-2831.

Bruun, S., J. W. Jensen, J. Magid, J. Lindedam and S. B. Engelsen (2010). "Prediction of the degradability and ash content of wheat straw from different cultivars using near infrared spectroscopy." *Industrial Crops and Products* **31**(2): 321-326.

Burns, D. A. and E. W. Ciurczak (2008). *Handbook of Near-Infrared Analysis - Third Edition*, Taylor & Francis Group, LLC.

Chadwick, D. T., K. P. McDonnell, L. P. Brennan, C. C. Fagan and C. D. Everard (2014). "Evaluation of infrared techniques for the assessment of biomass and biofuel quality parameters and conversion technology processes: A review." *Renewable and Sustainable Energy Reviews* **30**(0): 672-681.

Chen, H., C. Ferrari, M. Angiuli, J. Yao, C. Raspi and E. Bramanti (2010). "Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis." *Carbohydrate Polymers* **82**(3): 772-778.

Chong, I.-G. and C.-H. Jun (2005). "Performance of some variable selection methods when multicollinearity is present." *Chemometrics and Intelligent Laboratory Systems* **78**(1–2): 103-112.

Dare, P., J. Gifford, R. J. Hooper, A. H. Clemens, L. F. Damiano, D. Gong and T. W. Matheson (2001). "Combustion performance of biomass residue and purpose grown species." *Biomass and Bioenergy* **21**(4): 277-287.

Department of Chemistry and Biochemistry (2002). *Chapter 15: Infrared Spectroscopy: Theory*, in "Handbook for Organic Chemistry Lab", University of Colorado, Boulder.

Dessipri, E., E. Minopoulou, G. D. Chryssikos, V. Gionis, A. Paipetis and C. Panayiotou (2003). "Use of FT-NIR spectroscopy for on-line monitoring of formaldehyde-based resin synthesis." *European Polymer Journal* **39**(8): 1533-1540.

Dwivedi, Y. and S. B. Rai (2009). "Spectroscopic study of overtone and combination bands in aliphatic aldehydes." *Vibrational Spectroscopy* **49**(2): 278-283.

Emandi, A., C. Ileana Vasiliu, P. Budrugeac and I. Stamatin (2011). "Quantitative investigation of wood composition by integrated FT-IR and thermogravimetric methods." *Cellulose chemistry and technology.* **45**(9-10): 579-584.

Erol, M., H. Haykiri-Acma and S. Küçükbayrak (2010). "Calorific value estimation of biomass from their proximate analyses data." *Renewable Energy* **35**(1): 170-173.

Eskildsen, C. E. A. (2016). "Prediction of milk quality parameters using vibrational spectroscopy and chemometrics - opportunities and challenges in milk phenotyping ". *Ph.D. Thesis* University of Copenhagen, Faculty of Science

Esteves, B., A. Velez Marques, I. Domingos and H. Pereira (2013). "Chemical changes of heat treated pine and eucalypt wood monitored by FTIR." *Maderas. Ciencia y tecnología* **15**: 245-258.

European Union (2009). "Eurostat Statistics Energy, Energy ststistics – supply, transformation, consumption". http://epp.eurostat.ec.europa.eu.

Eurostat (Accessed: March 2015). "Share of renewables in energy consumption up to 15% in the EU in 2013".

Eurostat (Accessed: September 2017). "Energy, transport and environment indicators" http://ec.europa.eu/eurostat/documents/3217494/7052812/KS-DK-15-001-EN-N.pdf/eb9dc93d-8abe-4049-a901-1c7958005f5b.

Evans, P. A. (1991). "Differentiating “hard” from “soft” woods using Fourier transform infrared and Fourier transform spectroscopy." *Spectrochimica Acta Part A: Molecular Spectroscopy* **47**(9): 1441-1447.

Everard, C. D., K. P. McDonnell and C. C. Fagan (2012). "Prediction of biomass gross calorific values using visible and near infrared spectroscopy." *Biomass and Bioenergy* **45**(0): 203-211.

Fagan, C. C., C. Everard, C. P. O'Donnell, G. Downey, E. M. Sheehan, C. M. Delahunty and D. J. O'Callaghan (2007). "Evaluating Mid-infrared Spectroscopy as a New Technique for Predicting Sensory Texture Attributes of Processed Cheese." *Journal of Dairy Science* **90**(3): 1122-1132.

Fagan, C. C., C. D. Everard and K. McDonnell (2011). "Prediction of moisture, calorific value, ash and carbon content of two dedicated bioenergy crops using near-infrared spectroscopy." *Bioresource Technology* **102**(8): 5200-5206.

Faix, O. (1991). "Classification of Lignins from Different Botanical Origins by FT-IR Spectroscopy." *Holzforschung - International Journal of the Biology, Chemistry, Physics and Technology of Wood* **45**(s1): 21.

Feng, S., Z. Yuan, M. Leitch, H. Shui and C. C. Xu (2016). "Effects of bark extraction before liquefaction and liquid oil fractionation after liquefaction on bark-based phenol formaldehyde resoles." *Industrial Crops and Products* **84**: 330-336.

Feng, S., Z. Yuan, M. Leitch and C. C. Xu (2014). "Hydrothermal liquefaction of barks into bio-crude – Effects of species and ash content/composition." *Fuel* **116**: 214-220.

Filbakk, T., R. Jirjis, J. Nurmi and O. Høibø (2011). "The effect of bark content on quality parameters of Scots pine (Pinus sylvestris L.) pellets." *Biomass and Bioenergy* **35**(8): 3342-3349.

Gehrig, M., S. Pelz, D. Jaeger, G. Hofmeister, A. Groll, H. Thorwarth and W. Haslinger (2015). "Implementation of a firebed cooling device and its influence on emissions and combustion parameters at a residential wood pellet boiler." *Applied Energy* **159**: 310-316.

Gillespie, G. D., C. D. Everard and K. P. McDonnell (2015). "Prediction of biomass pellet quality indices using near infrared spectroscopy." *Energy* **80**(Supplement C): 582-588.

González, A., J.-R. Riba, R. Puig and P. Navarro (2015). "Review of micro- and small-scale technologies to produce electricity and heat from Mediterranean forests′ wood chips." *Renewable and Sustainable Energy Reviews* **43**(0): 143-155.

Grønli, M. G., G. Várhegyi and C. Di Blasi (2002). "Thermogravimetric Analysis and Devolatilization Kinetics of Wood." *Industrial & Engineering Chemistry Research* **41**(17): 4201-4208.

Hames, B. R., S. R. Thomas, A. D. Sluiter, C. J. Roth and D. W. Templeton (2003). "Rapid biomass analysis: New tools for compositional analysis of corn stover feedstocks and process intermediates from ethanol production." *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology* **105**(1-3): 5-16.

Harkin, J. M. and J. W. Rowe (1971). *Bark And Its Possible Uses.*, Research note FPL ; 091. Madison, WI: U.S. Dept. of Agriculture, Forest Service, Forest Products Laboratory.

Hindle, P. (2007). *Historical Development*, in "Handbook of Near-Infrared Analysis, Third Edition", CRC Press. 3-6.

Hotelling, H. (1933). *Analysis of a complex of statistical variables into principal components.* Baltimore, Warwick & York.

Huang, C., L. Han, Z. Yang and X. Liu (2009). "Ultimate analysis and heating value prediction of straw by near infrared spectroscopy." *Waste Management* **29**(6): 1793-1797.

Hurley, J. R. and R. B. Cattell (1962). "The procrustes program: Producing direct rotation to test a hypothesized factor structure." *Behavioral Science* **7**(2): 258-262.

Iroba, K. L. and Tabil (2015). *Photoacoustic Spectroscopy in the Assessment of the Quantitative Composition of the Biomass — Barley Straw*, in "Biofuels - Status and Perspective". K. Biernat. Rijeka, InTech. Ch. 15.

ISO (2014). Solid biofuels - Fuel specifications and classes - Part 1: General requirements, EN ISO 17225-1:2014.

ISO (2014). Solid biofuels - Fuel specifications and classes - Part 2: Graded wood pellets, EN ISO 17225-2:2014.

ISO (2017). Solid Biofuels - Sampling, EN ISO 18135:2017.

Jha, S. N. (2016). *Chapter 6 - Spectroscopy and Chemometrics*, in "Rapid Detection of Food Adulterants and Contaminants". San Diego, Academic Press. 147-214.

Kennard, R. W. and L. A. Stone (1969). "Computer Aided Design of Experiments." *Technometrics* **11**(1): 137-148.

Kortelainen, M., J. Jokiniemi, I. Nuutinen, T. Torvela, H. Lamberg, T. Karhunen, J. Tissari and O. Sippula (2015). "Ash behaviour and emission formation in a small-scale reciprocating-grate combustion reactor operated with wood chips, reed canary grass and barley straw." *Fuel* **143**(0): 80-88.

Lavine, B. (2003). "A user-friendly guide to multivariate calibration and classification, Tomas Naes, Tomas Isakson, Tom Fearn and Tony Davies, NIR Publications, Chichester, 2002, ISBN 0-9528666-2-5, £45.00." *Journal of Chemometrics* **17**(10): 571-572.

Lehtikangas, P. (2001). "Quality properties of pelletised sawdust, logging residues and bark." *Biomass and Bioenergy* **20**(5): 351-360.

Lestander, T. A., M. Finell, R. Samuelsson, M. Arshadi and M. Thyrel (2012). "Industrial scale biofuel pellet production from blends of unbarked softwood and hardwood stems—the effects of raw material composition and moisture content on pellet quality." *Fuel Processing Technology* **95**(0): 73-77.

Lestander, T. A., B. Johnsson and M. Grothage (2009). "NIR techniques create added values for the pellet and biofuel industry." *Bioresource Technology* **100**(4): 1589-1594.

Lestander, T. A. and C. Rhen (2005). "Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression." *Analyst* **130**(8): 1182-1189.

Liu, L., X. P. Ye, A. R. Womac and S. Sokhansanj (2010). "Variability of biomass chemical composition and rapid analysis using FT-NIR techniques." *Carbohydrate Polymers* **81**(4): 820-829.

Long, H., X. Li, H. Wang and J. Jia (2013). "Biomass resources and their bioenergy potential estimation: A review." *Renewable and Sustainable Energy Reviews* **26**: 344-352.

Lundqvist, J., A. Jacobs, M. Palm, G. Zacchi, O. Dahlman and H. Stålbrand (2003). "Characterization of galactoglucomannan extracted from spruce (Picea abies) by heat-fractionation at different conditions." *Carbohydrate Polymers* **51**(2): 203-211.

Lupoi, J. S., S. Singh, R. Parthasarathi, B. A. Simmons and R. J. Henry (2015). "Recent innovations in analytical methods for the qualitative and quantitative assessment of lignin." *Renewable and Sustainable Energy Reviews* **49**: 871-906.

Magelli, F., K. Boucher, H. T. Bi, S. Melin and A. Bonoli (2009). "An environmental impact assessment of exported wood pellets from Canada to Europe." *Biomass and Bioenergy* **33**(3): 434-441.

Manzone, M. (2016). "Quality, productivity, energy and costs of woodchip produced by Cedrus deodara plantations: A case study in Italy." *Biomass and Bioenergy* **92**: 81-87.

Martens, H., S. A. Jensen and P. Geladi (1983). "Multivariate linearity transformation for near infrared reflectance spectra of meat". Application Spectroscopy Proceedings of the Nordic Symposium, Applied Statistics., Stockholm Förlag Publication, Stavanger, Norway

Martens, H. and E. Stark (1991). "Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy." *J Pharm Biomed Anal* **9**(8): 625-635.

McCarthy, W. and G. Kemeny (2007). *Fourier Transform Spectrophotometers in the Near-Infrared*, in "Handbook of Near-Infrared Analysis, Third Edition", CRC Press. 79-91.

McKendry, P. (2002). "Energy production from biomass (part 1): overview of biomass." *Bioresource Technology* **83**(1): 37-46.

Minopoulou, E., E. Dessipri, G. D. Chryssikos, V. Gionis, A. Paipetis and C. Panayiotou (2003). "Use of NIR for structural characterization of urea–formaldehyde resins." *International Journal of Adhesion and Adhesives* **23**(6): 473-484.

Miranda, I., J. Gominho and H. Pereira (2012). "Incorporation of Bark and Tops in Eucalyptus Globulus Wood Pulping." **7**: 4350-4361.

Mola-Yudego, B., M. Selkimäki and J. R. González-Olabarria (2014). "Spatial analysis of the wood pellet production for energy in Europe." *Renewable Energy* **63**: 76-83.

Narodoslawsky, M. and I. Obernberger (1996). "From waste to raw material—the route from biomass to wood ash for cadmium and other heavy metals." *Journal of Hazardous Materials* **50**(2–3): 157-168.

Norgaard, L., A. Saudland, J. Wagner, J. P. Nielsen, L. Munck and S. B. Engelsen (2000). "Interval Partial Least-Squares Regression (iPLS): A Comparative

Chemometric Study with an Example from Near-Infrared Spectroscopy." *Applied Spectroscopy* **54**(3): 413-419.

Obernberger, I. and G. Thek (2010). *The Pellet Handbook. The Production and Thermal Utilization of Pellets*, Earthscan, London - Washington DC.

Pandey, K. K. and A. J. Pitman (2003). "FTIR studies of the changes in wood chemistry following decay by brown-rot and white-rot fungi." *International Biodeterioration & Biodegradation* **52**(3): 151-160.

Pasquini, C. (2003). "Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications." *Journal of the Brazilian Chemical Society* **14**: 198-219.

Pavia, D. L., G. M. Lampman and G. S. Kriz (2001). *Infrared Spectroscopy*, in "Introduction to spectroscopy: a guide for students of organic chemistry", Fort Worth: Harcourt College Publishers.

Pearson, K. (1901). "On lines and planes of closest fit to systems of points in space." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**(11): 559-572.

Pecoraro, E., B. Pizzo, A. Alves, N. Macchioni and J. C. Rodrigues (2015). "Measuring the chemical composition of waterlogged decayed wood by near infrared spectroscopy." *Microchemical Journal* **122**: 176-188.

Popescu, C.-M. and M.-C. Popescu (2013). "A near infrared spectroscopic study of the structural modifications of lime (Tilia cordata Mill.) wood during hydro-thermal treatment." *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **115**: 227-233.

Popescu, C.-M., G. Singurel, M.-C. Popescu, C. Vasile, D. S. Argyropoulos and S. Willför (2009). "Vibrational spectroscopy and X-ray diffraction methods to establish the differences between hardwood and softwood." *Carbohydrate Polymers* **77**(4): 851-857.

Posom, J., A. Shrestha, W. Saechua and P. Sirisomboon (2016). "Rapid non-destructive evaluation of moisture content and higher heating value of Leucaena leucocephala pellets using near infrared spectroscopy." *Energy* **107**: 464-472.

Proskurina, S., J. Heinimö, M. Mikkilä and E. Vakkilainen (2015). "The wood pellet business in Russia with the role of North-West Russian regions: present trends and future challenges." *Renewable and Sustainable Energy Reviews* **51**: 730-740.

Purohit, P., A. K. Tripathi and T. C. Kandpal (2006). "Energetics of coal substitution by briquettes of agricultural residues." *Energy* **31**(8–9): 1321-1331.

Rajalahti, T., R. Arneberg, F. S. Berven, K.-M. Myhr, R. J. Ulvik and O. M. Kvalheim (2009). "Biomarker discovery in mass spectral profiles by means of selectivity ratio plot." *Chemometrics and Intelligent Laboratory Systems* **95**(1): 35-48.

Reyes, P., R. Teixeira Mendonça, J. Rodríguez, P. Fardim And B. Vega (2013). "Characterization of the hemicellulosic fraction obtained after pre-hydrolysis of pinus radiata wood chips with hot-water at different initial PH." *Journal of the Chilean Chemical Society* **58**: 1614-1618.

Rinnan, Å., M. Andersson, C. Ridder and S. B. Engelsen (2014). "Recursive weighted partial least squares (rPLS): an efficient variable selection method using PLS." *Journal of Chemometrics* **28**(5): 439-447.

Rinnan, Å., F. v. d. Berg and S. B. Engelsen (2009). "Review of the most common pre-processing techniques for near-infrared spectra." *TrAC Trends in Analytical Chemistry* **28**(10): 1201-1222.

Russ, A. and M. Fiserova (2011). "Estimation of hardwood species in mixture by near infrared spectroscopy." *Wood Res Slovak* **56**: 93-104.

Sandak, A., J. Sandak, M. Zborowska and W. Prądzyński (2010). "Near infrared spectroscopy as a tool for archaeological wood characterization." *Journal of Archaeological Science* **37**(9): 2093-2101.

Santoni, I., E. Callone, A. Sandak, J. Sandak and S. Dirè (2015). "Solid state NMR and IR characterization of wood polymer structure in relation to tree provenance." *Carbohydrate Polymers* **117**: 710-721.

Savitzky, A. and M. J. E. Golay (1964). "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." *Analytical Chemistry* **36**(8): 1627-1639.

Schwanninger, M., J. Rodrigues and K. Fackler (2011). "A review of band assignments in near infrared spectra of wood and wood components." *Journal of Near Infrared Spectroscopy* **19**(5): 287-308.

Schwanninger, M., J. C. Rodrigues, H. Pereira and B. Hinterstoisser (2004). "Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose." *Vibrational Spectroscopy* **36**(1): 23-40.

Selkimäki, M., B. Mola-Yudego, D. Röser, R. Prinz and L. Sikanen (2010). "Present and future trends in pellet markets, raw materials, and supply logistics in Sweden and Finland." *Renewable and Sustainable Energy Reviews* **14**(9): 3068-3075.

Siesler, H. (2007). *Basic Principles of Near-Infrared Spectroscopy*, in "Handbook of Near-Infrared Analysis, Third Edition", CRC Press. 7-19.

Sikkema, R., M. Steiner, M. Junginger, W. Hiegl, M. T. Hansen and A. Faaij (2011). "The European wood pellet markets: current status and prospects for 2020." *Biofuels, Bioproducts and Biorefining* **5**(3): 250-278.

Simon, M., Y. Brostaux, C. Vanderghem, B. Jourez, M. Paquot and A. Richel (2014). "Optimization of a formic/acetic acid delignification treatment on beech wood and its influence on the structural characteristics of the extracted lignins." *Journal of Chemical Technology & Biotechnology* **89**(1): 128-136.

Simonović, J., J. Stevanic, D. Djikanović, L. Salmén and K. Radotić (2011). "Anisotropy of cell wall polymers in branches of hardwood and softwood: a polarized FTIR study." *Cellulose* **18**(6): 1433-1440.

Sjöström, E. (1993). *Wood Chemistry (Second Edition).* San Diego, Academic Press.

Sluiter, J. B., R. O. Ruiz, C. J. Scarlata, A. D. Sluiter and D. W. Templeton (2010). "Compositional Analysis of Lignocellulosic Feedstocks. 1. Review and Description of Methods." *Journal of Agricultural and Food Chemistry* **58**(16): 9043-9053.

Spinelli, R., L. Ivorra, N. Magagnotti and G. Picchi (2011). "Performance of a mobile mechanical screen to improve the commercial quality of wood chips for energy." *Bioresource Technology* **102**(15): 7366-7370.

Spinelli, R., C. Nati, L. Sozzi, N. Magagnotti and G. Picchi (2011). "Physical characterization of commercial woodchips on the Italian energy market." *Fuel* **90**(6): 2198-2202.

Stark, N. M., Z. Cai and C. Carll (2010). *Wood-based composite materials: panel products, glued-laminated timber, structural composite lumber, and wood-nonwood composite materials: chapter 11*, in "Wood handbook: wood as an engineering material". Centennial ed. General technical report FPL; GTR-190. Madison, WI: U.S. Dept. of Agriculture, Forest Service, Forest Products Laboratory. 11.11-11.28.

Stuart, B. H. (2005). *Inorganic Molecules*, in "Infrared Spectroscopy: Fundamentals and Applications", John Wiley & Sons, Ltd. 95-111.

Sun, L. and K. Niquidet (2017). "Elasticity of import demand for wood pellets by the European Union." *Forest Policy and Economics* **81**(Supplement C): 83-87.

Tomlinson, S. K., O. R. Ghita, R. M. Hooper and K. E. Evans (2006). "The use of near-infrared spectroscopy for the cure monitoring of an ethyl cyanoacrylate adhesive." *Vibrational Spectroscopy* **40**(1): 133-141.

Tooyserkani, Z., S. Sokhansanj, X. Bi, J. Lim, A. Lau, J. Saddler, L. Kumar, P. S. Lam and S. Melin (2013). "Steam treatment of four softwood species and bark to produce torrefied wood." *Applied Energy* **103**: 514-521.

Toscano, G., D. Duca, A. Amato and A. Pizzi (2014). "Emission from realistic utilization of wood pellet stove." *Energy* **68**(0): 644-650.

Toscano, G., D. Duca, E. Foppa Pedretti, A. Pizzi, G. Rossini, C. Mengarelli and M. Mancini (2016). "Investigation of woodchip quality: Relationship between the most important chemical and physical parameters." *Energy* **106**: 38-44.

Toscano, G., G. Riva, E. Foppa Pedretti, F. Corinaldesi, C. Mengarelli and D. Duca (2013). "Investigation on wood pellet quality and relationship between ash content and the most important chemical elements." *Biomass and Bioenergy* **56**: 317-322.

Tsuchikawa, S. and H. Kobori (2015). "A review of recent application of near infrared spectroscopy to wood science and technology." *Journal of Wood Science* **61**(3): 213-220.

Tumuluru, J. S., S. Sokhansanj, J. Hess, C. Wright and R. Boardman (2011). "A Review on Biomass Torrefaction Process and Product Properties for Energy Applications." *Ind Biotechnol* **7**(5): 384-401.

Vassilev, S. V., D. Baxter, L. K. Andersen and C. G. Vassileva (2010). "An overview of the chemical composition of biomass." *Fuel* **89**(5): 913-933.

Vassilev, S. V., C. G. Vassileva, Y.-C. Song, W.-Y. Li and J. Feng (2017). "Ash contents and ash-forming elements of biomass and their significance for solid biofuel combustion." *Fuel* **208**(Supplement C): 377-409.

Weiss, W. P. (2017). "A 100-Year Review: From ascorbic acid to zinc—Mineral and vitamin nutrition of dairy cows." *Journal of Dairy Science* **100**(12): 10045-10060.

Werkelin, J., B.-J. Skrifvars, M. Zevenhoven, B. Holmbom and M. Hupa (2010). "Chemical forms of ash-forming elements in woody biomass fuels." *Fuel* **89**(2): 481-493.

Williams, P. C. (2001). *Implementation of near-infrared technology*, in "Near Infrared Technology in the Agriculture and Food Industries", American Association of Cereal Chemists, Minnesota, USA.

Wold, S. and M. SjÖStrÖM (1977). *SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy*, in "Chemometrics: Theory and Application", AMERICAN CHEMICAL SOCIETY. **52**. 243-282.

Wold, S., M. Sjöström and L. Eriksson (2001). "PLS-regression: a basic tool of chemometrics." *Chemometrics and Intelligent Laboratory Systems* **58**(2): 109-130.

Xu, F., J. Yu, T. Tesso, F. Dowell and D. Wang (2013). "Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review." *Applied Energy* **104**(0): 801-809.

Yan, J., N. Villarreal and B. Xu (2013). "Characterization of Degradation of Cotton Cellulosic Fibers Through Near Infrared Spectroscopy." *Journal of Polymers and the Environment* **21**(4): 902-909.

Yang, Z., B. Lu, A. M. Huang, Y. N. Liu and X. Q. Xie (2012). "Rapid identification of softwood and hardwood by near infrared spectroscopy of cross-sectional surfaces." *Spectrosc Spect Anal* **32**(7): 1785-1789.

Yildiz, G., F. Ronsse, R. Venderbosch, R. v. Duren, S. R. A. Kersten and W. Prins (2015). "Effect of biomass ash in catalytic fast pyrolysis of pine wood." *Applied Catalysis B: Environmental* **168–169**: 203-211.

Zhou, C., W. Jiang, Q. Cheng and B. K. Via (2015). "Multivariate Calibration and Model Integrity for Wood Chemistry Using Fourier Transform Infrared Spectroscopy." *Journal of Analytical Methods in Chemistry* **2015**: 9.

Zhou, C., W. Jiang, B. K. Via, O. Fasina and G. Han (2015). "Prediction of mixed hardwood lignin and carbohydrate content using ATR-FTIR and FT-NIR." *Carbohydrate Polymers* **121**: 336-341.