NIRWOOD: an EU Innovation Project to Determine Species and Geographical Origin of Timber Using NIR Spectrometry

Sierra, V.[1]; De Luque, M.[2]; Álvarez-San Millán, M.[1]; Pérez, S.[2]; García, E.[1]; Luque, L.[2]
[1] Spectrapply, Polígono Industrial Puente Nora, Calle E, 2, 2ª planta, 33420 Lugones, Asturias, Spain, www.spectrapply.com
[2] GEA Forestal, Tercio, 20, Bajo D, 28019 Madrid, Spain, www.geaforestal.com
[*]Corresponding author: i+d@geaforestal.com

ABSTRACT

Near Infrared Spectrometry (NIR) will be a revolutionary innovation for the wood sector. The ability of this analytical technique to determine chemical and physical properties of wood has been positively tested. But NIR has also been proved to be an useful tool to determine other indirect data such as the botanical species or its geographical origin, because environmental features give a unique fingerprint to any timber sample. To extract this information from infrared spectra it is necessary to carry out a complex statistical treatment (chemometrics) and to compare every new wood sample with a representative database of wood spectra. To achieve this goal, the Spanish companies GEA Forestal and Spectrapply have joined together in a project partially funded by the European Union (Horizon 2020 Program) called NIRWOOD. In this work we show our first results carried out with three commercial spectrometers. The results obtained from timber of 16 commercial species in 25 localities in Spain are evaluated, reaching percentages of correct identification always over 80%, as well as discovering key methodological questions. NIR technology could be very useful to improve the wood traceability and to comply with the European regulations on timber trade.

Key words: EUTR, Near-infrared, Traceability, Wood

1. INTRODUCTION

Wood from illegal logging accounts for around 15% of the total imports in the European Union. Illegal logging causes significant global economic losses (around 15 billion $/year, according to the World Bank), and heavy environmental and social damages. The European Union regulation on timber trade (EUTR, since 2010) aims to tackle the illegal timber entering its territory. But nowadays, the timber traceability methods mainly consist in documentary control, that can be easily falsified, and more rarely slow and expensive laboratory analyses of DNA or Isotopes. Nirwood is an innovation project, partially financed by the European Union through its program Horizon 2020, SME Instrument that will incorporate Near Infrared (NIR) Spectrometry as a technological solution to determine botanical species and geographical origin of the wood in a simple, fast and affordable way. The aim of the project is to create a huge reference wood database of those species more commonly commercialized in the European Union and to develop a NIR device adapted to timber analysis.

NIR spectroscopy analyses the absorption or transmission of light in the range of the electromagnetic spectrum of the near infrared radiation, from 780 to 2500 nm. Depending on the chemical composition and physical structure of each analysed material, a specific absorbance spectrum will be obtained, and it will behave as the "fingerprint" of the material. The information behind the spectra directly obtained by the spectrometer has to be extracted through complex chemometrical treatments. NIR spectrometry applied to timber provides information about chemical composition and structure (moisture, lignin content, extractive content, sugars, etc.), and physical and mechanical properties (strength, density, etc.). There are also a number of works testing NIR technology on solid wood, leaves, needles or charcoal that demonstrate the great potential of this technique to discriminate between botanical species. NIR spectrometry is also sensitive enough to determine phenotypical differences in timber due to the influence of local conditions in the tree growing places (soil, altitude, climate, silvicultural treatments, and others) allowing to group specimens by its geographical origin. For more information about these applications, see the special issues dedicated to wood and derivatives of the Journal of Near Infrared Spectroscopy of 2010 and 2011 and other recent reviews such as Tsuchikawa and Kobori (2015) or Tsuchikawa and Schwanninger (2013). Some authors already have noticed the potential of this technology to fight against illegal logging (Pastore *et al*., 2011; Sandak *et al.*, 2011; Dormontt *et al*., 2015; Nisgoski *et al*., 2016).

In this work we show the results of our preliminary sampling session of timber carried out in 25 Spanish sawmills working with well controlled timber species and its origin. The aim of this work is to test three commercial NIR devices with the aim of determine the most appropriate sampling methodology, to check their ability to discriminate both botanical species and geographical origin and to determine the minimum specifications needed to develop the Nirwood prototype before to start the compilation of a representative international reference database.

2. METHODOLOGY

Sampling has focused in eight of the most commercial species in Spain, but also some tropical species to introduce some variability. Spectra from a total of 16 species have been collected in 25 sawmills at two different ecoregions (Atlantic and Mediterranean). We have worked with small local sawmills to ensure that the timber was obtained within a maximum radius of 100 km. We have taken 69 pairs species/origin: 8 temperate species coming from 54 localities (22 provinces) and eight tropical species coming from 10 different sources (taken the importing sawmill as an origin due to the uncertainty of the real timber source). We have taken a minimum of 20 samples per pair species/locality. We took samples in both logs and boards and both in tangential and transversal sections. In some cases, we previously sanded the wood. The species are: *Castanea sativa*, *Eucaliptus globulus*, *Eucaliptus nitens*, *Pinus nigra*, *Pinus pinaster*, *Pinus radiata*, *Pinus sylvestris*, *Quercus robur*, *Chlorophora excelsa* (Iroko), *Entandrophragma cylindricum* (Sapelli), *Didelotia africana* (Gombé), *Cedrella odorata* (Cedar of Brazil), *Triplochiton scleroxylon* (Ayous), *Millettia laurentii* (Wengue), *Erythrophleum ivorense* (Elondo), and *Hymenaea courbaril* (Jatoba).

2.1. NIR Devices

Three commercial devices were used for this preliminary study: LABSPEC®PRO (Analytical Spectral Device-ASD) working in reflectance mode from 350-2500 nm (1nm spectral interval); the handheld NIR spectrometer MICROPHAZIR (Thermofisher) that measures in reflectance mode form 1600-2400nm range (pixel resolution 8 nm, optical resolution 12 nm); and the portable spectrophotometer MICRONIR 1700 (VIAVI) that measures in reflectance from 950 to 1650nm (spectral resolution of 10 nm).

2.2. Chemometrics

Identification of wood or any other material by spectral data depends on classification and pattern recognition techniques. The Unscrambler X chemometric program (version 10.4, from CAMO Software AS) was used to analyse the data. Spectral data from each device were averaged, analysed and outliers were removed. The final spectral population was divided by random selection in two sets: the calibration data set (training set) and the validation data set.

2.2.1. Pre-processing

Pre-processing eliminates noise and removes physical phenomena in the spectra. The most commonly used pre-treatments were: Standard Normal Variate (SNV), Detrend (D); multiplicative Scatter correction (MSC), Savitzky-Golay first order derivative (SG1), Savitzky-Golay second order derivative (SG2) and different combinations of them.

2.2.2. Exploratory Analysis

Principal Component Analysis (PCA) is an unsupervised pattern recognition technique. The computation of PC's for the PCAs was developed using nonlinear iterative partial least squares (NIPALS) algorithm, over the training set data. The optimal number of PC's was determined based on the cumulative variance percent of PC's.

2.2.3. Pattern Recognition Methods

Individual models were based on the NIPALS algorithm and validated with cross validation for model development. SIMCA, PCA-LDA and PLS-DA classifications were applied over the row (without any treatment) training set and also over the training set after being transformed with different pre-treatments such as, first (polynomial order = 1, smoothing point = 4) and second (polynomial order = 2, smoothing point = 5) derivatives of Savitzky-Golay and Standard Normal Variate and Detrend (SNVD) and Multiplicative Scatter Correction (MSC) to obtain predictive models. All the different models created were used to classify samples of the validation data set. Recognition and prediction ability of the models were evaluated by the percentage of samples from the validation data set (rate prediction) classified correctly on their corresponding category.

3. RESULTS

3.1. Timber sections

In a previous work (De Luque *et al*., 2017) we already observed differences in absorbance depending on the orientation of the fibres, being greater in tangential and radial sections than in transversal ones. In this case, spectral signatures of roundwood and sawnwood from the same species (*Quercus robur*) were tested again, showing clear spectral differences between them. For this reason, we consider that they should be treated independently for modelling development. In addition, the edges of logs and boards of tropical timber are usually labelled with paint and treated with parfine which prevent sampling on it.

3.2. Sapwood and heartwood

Our results showed a clear spectral difference between the sapwood and the heartwood, showing sapwood higher dispersion and lower repeatability. For this reason, the main part of the sampling was done on the heartwood.

3.3. Surface preparation

Sampling before and after hand-sanding the surface of the transversal sections of sapelli boards showed surprising results, because dispersion of the samples was bigger after manual sanding than before. It is probably due changes in the intensity of the sanding by the worker.

3.4. Samples classification

Comparison between pattern recognition methods after pre-processing:

- Discrimination between temperate *vs* tropical species (boards, transverse sections): the best results for all three devices were obtained by PLS-DA (100% correct classification).

- Discrimination between temperate species (*Pinus pinaster*, *Pinus radiata* and *Quercus robur*, boards transverse sections) tested on boards gave good results in PCA analysis (85% of the cumulative variance explained by the two first PC's). It allowed us to use classification methods based on PC's, and then LDA, SIMCA and PLS-DA methods were performed. The best results for all the devices were obtained by PLS-DA models were all the samples of the validation data set were correctly classified, but only in the case of LABSPEC all the models (LDA, PLS-DA and SIMCA) achieved the 100% of correct classification.

- Discrimination about origin of the samples (5 pairs species/origin, boards and transverse sections). Again, the good results of PCA allows to use LDA, SIMCA and PLS-DA methods and again PLS-DA allowed correct classifications by category (pair specie/origin) of all the validation samples. LABSPEC was the device that showed the best results (100% of correct classification) but MICRONIR showed also very good percentages over 90%.

- Discrimination between tropical species (boards, transverse sections). Exploratory analysis showed well defined groups of spectral data depending on the species and the best results were obtained by PLS-DA models with percentages of correct classification over 90%. The best results were again achieved with LABSPEC models that reached 95% of correct specie identification. SIMCA did not report good results.

- Discrimination between sources of tropical species (boards, transverse sections). Again, good results (over 93%) were reported for the three devices in the case of LDA and PLS-DA models. Results for SIMCA were poorer.

- Discrimination between temperate species (70 pairs species/origin) in logs (transverse sections). Results for LDA and PLS-DA showed a percentage of success in species classification over 77%, independently of the NIR device considered. This species discrimination was slightly lower than the ones reported for sawnwood (which were over 90%). SIMCA models did not reported good results for species classification of temperate roundwood for any of the commercial devices, showing low percentages of success, between 45 and 53%. It is important to note that percentages of correct classification differ depending on the species that have been tested. We could see how the percentages were higher (over 80%) for *C. sativa*, *P. pinaster*, *P. radiata* and *P. sylvestris*, independent of the device used, and lower percentages of correct classification were found for *E. globulus*, *E. nitens*, and *P. nigra* with percentages that in some cases decreased to 50% (random classification).

4. DISCUSSION

The good percentages of correct classification obtained in all the cases during this sampling test, confirm the great potential of NIR technology as a useful tool to discriminate between timber botanical species and geographical origin both in temperate and tropical species, both in sawnwood and logs, and both in transversal and tangential/radial sections. We also confirmed that 20 samples are sufficient to characterize a pair species/origin, and that 5 spots per sample makes the sampling statistically robust.

Due to natural weathering, decay and chemical changes are common processes on the timber surface. Other authors suggest the utility of previous cleaning and sanding of the samples, extracting core samples, cutting out slices from logs or collecting chain-saw chips (Cooper *et al*., 2011; Sandak *et al*., 2011). However, while desirable, after taking into consideration a number of factors (such as simplicity, low cost and fast *in situ* sampling) and looking at our good discrimination results on raw surfaces without chemically treated sections (paint, or parafine), we consider acceptable to work on raw materials in our future sampling campaigns, but always avoiding especially dirty areas or resin spots on the surface.

Despite the good results obtained after sampling in transverse sections of logs and boards (results exposed in this work), the better absorbance of tangential and radial sections lead us to consider these longitudinal sections as a better source of information for the future sampling campaigns. In any case, models must be built independently. Some authors have found the same influence of the timber face in the quality of the results (Braga *et al.*, 2011; Cooper *et al.*, 2011) while others didn't find significative differences (Da Silva *et al.*, 2016).

One of the objectives of this test was to determine the most useful NIR wavelength range to characterize wood. Schwanninger *et al.* (2011) already did a band assignment of wood components, emphasizing the importance of the underlying chemistry. Our best correlations were obtained using the whole NIR spectral range (LABSPEC), as Adedipe *et al.* (2008) did when comparing similar devices. Other authors have found more useful spectral regions to distinguish species. Some examples are: 1400-2500nm (Asner *et al.*, 2014), visible region (Belmonte *et al.*, 2014), 1000-1600nm (Bolzón de Muñiz *et al.*, 2016), and 1850-2500nm (Espinoza *et al.*, 2012). Nisgoski *et al*, (2016) point that this most useful range varies depending on the species. In our case, even obtaining the best results with LABSPEC device (350-2500 nm), we consider that both MICRONIR (950-1650nm) and specially MICROPHAZIR (1600-2400nm), covering less than the half of the NIR spectral range, offered enough good results of correct classification. That is the case of similar tests carried out by Da Silva *et al.* (2016).

Heartwood and sapwood have different structural and chemical composition and these differences can be recorded by NIR spectrometers. Our most homogeneous results were those taken on heartwood, especially in the case of coniferous samples. For this reason we consider it is better to take samples not far from the core of the log or its equivalent in boards.

About chemometrics, it looks clear that PCA analysis gives good results and it allowed us to use classification methods based on PC's: LDA, SIMCA and PLS-DA. LDA. PLS-DA model has always produced the highest percentages of correct classification, usually over 90%, while SIMCA results were poorer. PLS-DA method has been widely used by other researchers to discriminate between species and origins with positive results (Braga *et al.*, 2011; Cooper *et al.*, 2011; Pastore *et al*,. 2011, Prades *et al.*, 2013; Bergo *et al.*, 2016). SIMCA model has been successfully used by other researchers (p.e. Adedipe *et al.*, 2008; Bolzón de Muñiz *et al.*, 2016) but it gave us worst results. Other authors did correct discrimination using just PCA pattern recognition method (p.e. Nisgoski *et al.*, 2016).

Our results show different percentages of correct classification depending on the species. It has also been observed by other authors (Espinoza *et al.*, 2012) and could be attributed to different genetical relationship between species or to methodological questions (p.e. the different number of samples taken for each one of them). In this sense, Asner *et al.* (2014) pointed that the spectra were dominated more by phylogeny than by abiotic filtering.

5. CONCLUSIONS

This work allowed us to clarify some of the requirements that have to be taken into consideration for the development of the NIRWOOD project (sampling collection, NIR device and chemometrics). Results from raw surfaces are good enough to discriminate species and geographical origins. Tangential/radial sections give better results than transverse ones, so sawnwood and roundwood have to be treated as independent populations for modelling purposes. Heartwood spectra are more homogeneous than sapwood ones. Correct discrimination is higher in boards than in logs. A minimum of 20 samples (individual logs or boards) per pair specie/origin have to be measured to properly describe each category. The best results for classification in validation were normally obtained by PLS-DA models, LDA models reported were also acceptable while SIMCA showed a problem of overlapping between species in most cases. Full NIR range gives the most reliable results but partial ranges also offered enough good results usually above 90% (particularly 1600-2400nm range) to make a correct classification.

Acknowledgements: This work has been partially funded by the EU, NIRWOOD Project. We appreciate the support of the selfless people who allowed us to sample in these 25 sawmills.

REFERENCES

Asner, G.P.; Martin, R.E.; Carranza-Jiménez, L.; Sinca, F.; Tupayachi, R.; Anderson, C.B.; Martinez, P. (2014): *Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region.* New Phytologist 204: pp. 127–139.

Belmonte, D.; Paim. A.C.; Soares, R.; Sabadi, M.; Mesquita, M.; Kervald, L. (2014) *Spectroradiometry of commercial wood veneers in the visible and near infrared spectra*. Floresta, Curitibia 44 (1): pp. 125-132.

Bergo, M.C.J.; Pastore, T.C.M.; Coradin, V.T.R.; Wiedenhoeft, A.C.; Braga, J.W.B. (2016): *NIRS Identification of Swietenia macrophylla is robust across specimens from 27 countries.* IAWA Journal, 37 (3): pp. 420-430.

Bolzon de Muñiz, G.E.; Carneiro, M.E.; Rodrigues, F.; Zatt, F.; Nisgoski, S. (2016): *Wood and charcoal identification of five species from the miscellaneous group known in Brazil as "Angelim" by Near-IR and wood anatomy.* Maderas. Ciencia y tecnología 18 (3): pp. 505-522.

Braga, J.W.B.; Monteiro Pastore, T.C.; Rauber, V.T.; Alves, J.A.; Ribeiro da Silva, A. (2011): *The use of Near Infrared spectroscopy to identify solid wood specimens of* Swietenia macrophylla *(CITES Appendix II)*. IAWA Journal, Vol. 32 (2): pp. 285–296.

Cooper, P.A.; Jeremic, D.; Radivojevic, S.; Ung, Y.T.; Leblon, B. (2011): *Potential of near-infrared spectroscopy to characterize wood products*. Can. J. For. Res. 41: pp. 2150–2157.

Da Silva, D.C.; Braga, J.W.C.; Soares, L.F.; Bergo, M.C.J.; Teixeira, N.C.; Gontijo, A.B.; Da Costa, N.R.; Coradin, V.T.R.; Wiedenhoeft, A.C.; Pastore, T.C.M. (2016) *Identification of the origin of Big-leaf mahogany Word from five Latin American countries via NIRS and PLS-DA*. 2nd Workshop "NIR spectroscopy for wood science and technology research", Italy. pp. 28-30.

De Luque, M.; Álvarez San Millám, M.; Pérez, S.; Sierra, V.; Luque, L.; García, E. (2017): *Desarrollo de una aplicación basada en la Espectroscopia de Infrarrojo Cercano (NIR) para la identificación de la especie y el origen geográfico de la madera.* VII Congreso Forestal Español, Plasencia, Spain.

Dormontt, E.E; Boner, M.; Braun, B.; Breulmann, G.; Degen, B.; Espinoza, E.; Gardner, S.; Guillery, P.; Hermanson, M.J.C.; Koch, G.; Lee, S.L.; Rimbawanto, M.A.; Thomas, D.; Wiedemhoeft, A.C.; Yin, Y.; Zahnen, J.; Lote, A.J. (2015): *Forensic timber identification: It's time to integrate disciplines to combat illegal logging*. Biol Cons. 191: pp. 790-798.

Espinoza J. A:, Hodge, G. R. and Dvorak, S. 2012. *The potential use of near infrared spectroscopy to discriminate between different pine species and their hybrids.* J. Near. Infrared Spectrosc., 18: 437-447.

Pastore, T.C.M.; Braga, J.W.B.; Coradin; V.T.R.; Esteves, W.L.; Arakaki, E.Y.; Alves, J.A.; Bolzón de Muñiz, G.I.; Bressan, O.A.; Davrieux, F. (2011): *Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupix.* Holzforschung 65: pp. 73-80.

Nisgoski, S.; Schardosin, F.Z.; Batista, F.R.R.; Bolzón de Muñiz, G.I.; Carneiro, M.E. (2016) : *Potential use of NIR spectroscopy to identify* Cryptomeria japonica *varieties from southern Brazil*. Wood Sci. Technol. 50: pp. 71-80.

Prades López, C.; Gómez, I.; García Olmo, J.; González Adrados, J.R. (2013): *Análisis discriminante de la procedencia geográfica de planchas y tapones de corcho natural mediante espectroscopia en el infrarrojo cercano.* VIº Congreso Forestal Español, Vitoria. pp. 1-13.

Sandak, A.; Sandak, J.; Negri, M. (2011): R*elationship between near-infrared (NIR) spectra and the geographical provenance of timber.* Wood Sci. Technol. 45 (1): pp. 35-48.

Schwanninger, M.; Rodrigues, J.C.; Fackler, K. (2011): *A review of band assignments in near infrared spectra of wood and wood components.* J. Near Infrared Spectr. 19: pp. 287-308.

Tsuchikawa, S.; Schwanninger, M. (2013): *A review of recent near-infrared research for wood and paper (Part 2).* Applied Spectroscopy Reviews 48: pp. 560-587.

Tsuchikawa, S.; Kobori, H. (2015): *A review of recent application of near infrared spectroscopy to wood science and technology.* J. Wood Sci. 61: pp. 213-220.