WOOD RESEARCH
56 (1): 2011
93-104

ESTIMATION OF HARDWOOD SPECIES IN MIXTURE BY NEAR INFRARED SPECTROSCOPY

Albert Russ, Mária Fišerová
Pulp And Paper Research Institute, Bratislava, Slovak Republic

(Received November 2010)

ABSTRACT

PCA analyses of hardwood NIR spectra showed high separation capabilities between ten various hardwood species. A series of measurements in various combinations were made with emphasis on standards and unknown samples drawn from the same lot and at around the same time span. We explain this approach to have a profound effect on the overall selectivity gain. Separation of hardwood species based on their density and classifying them according to the first PCA model led to classification success rate of 96.5 %. Separation of all hardwood species into independent components led to lower success rate of 79 %. Still, due to switched misidentification on the long run, only the total counts matter, not every single independent identification. Considering this, the selectivity success rises to 85 %.

KEYWORDS: NIR spectroscopy, wood species, wood chips, discrimintant analysis, hardwood, PCA.

INTRODUCTION

Hardwoods has been of interest as a papermaking raw materials for many years, which is related with an increasing demand of hardwood kraft pulps for printing papers. The papermaking properties of hardwood kraft pulps are primarily determined by wood species used as a raw material. Therefore a fast and reliable qualitative and quantitative determination of wood mixture composition entering the pulping process is of utmost importance for optimisation of production.

Near infrared (NIR) technology in terms of transmitted or reflected spectra of electromagnetic waves ranging from 800 to 2500 nm has been mainly used for non-destructive measurements of organic materials such as agricultural products and foods. However, it shows great potential in pulp and paper industry and in the case of wood or forest products, NIR spectroscopy should be widely used. Application of NIR spectroscopy in wood and pulp research was thoroughly reviewed (Tsuchikawa 2007).

As wood is an organic chemical complex, the interpretation of NIR spectrum is complicated. However, some researchers tried to clarify such information-rich spectroscopic signals. The origins of bands in the NIR spectra of eucalyptus globules wood were investigated (Michell and Schimleck

1996). Many partial correlations were found showing a high degree of inter correlation between bands and confirming that NIR bands arose from combinations of several fundamental bands. Furthermore, correlations were found between bands in NIR spectra and chemical composition of eleven eucalypts wood samples (Schimleck et al. 1997). NIR spectroscopy was used for investigation a large number of mixed species that displayed extremely wide variations in wood chemical composition, anatomy, and physical properties (Schimleck et al. 2001).

Wood species can be also determined by NIR reflectance spectroscopy evaluated with Chemometrics methods such Partial least squares projection to latent structures (PLS) and Principal component analysis (PCA. PCA analyses of NIR spectra of standards for hardwood species have been carried out with a varying selectivity among hardwood species (Russ et al. 2009). Selectivity between most of the hardwood species were around 60-70 % and with the exception of black locust, selectivity was unremarkable and suffering from even greater selectivity loss if many (more than three or four) hardwood species were included in a single PCA analysis.

PCA analysis of twelve hardwood species NIR spectra showed high separation capabilities between various hardwood species analyzed in pairs and triplets (Russ et al. 2010).

The objective of this study was to verify the possibility of wood species determination in a multi component mixture of hardwood chips by NIR spectroscopy. The results will be used in development of an on-line method for determination of wood species entering the digester for pulp production.

MATERIAL AND METHODS

Material

A total of ten different species of wood samples in form of chips were tested: beech (*Fagus silvatica L.*), oak (*Quercus robur L.*), turkey oak (*Quercus cerris L.*), black locust (*Robinia pseudoacacia* L.), hornbeam (*Carpinus betulus L.*), ash (*Fraxinus excelsior L.*), poplar (*Populus alba L.*), alder (*Alnus glutinosa L.*), birch (*Betula alba L.*), aspen (*Populus tremula L.*) and sweet cherry (*Cerasus avium L.*). For NIR measurements ten chips for each wood species were selected as standards for discriminant analysis calibration, based on two criteria. Firstly, the measured side of chips was selected as uniform as possible. Secondly, chips exhibiting a range of shades within a single wood species were selected such that they would represent this range of shades. Therefore, we tried to handpick chips with similar frequency of occurrence as in the others.

Wood chips of the listed species of hardwoods were used for analysis and were processed in the following way:

- Wood chips of varying moisture between 26 % and 46 % depending on the wood species were stored in polyethylene bags in a refrigerator for 4-5 days before analyses. Refrigeration prevented decaying processes that would greatly influence the NIR spectra outcome.
- Prior to analysis , the wood chips were taken out of the bags and those with uniform surface area were selected. Before the measurements, all chips were adjusted to room temperature for an approximate period of one hour. This step eliminates water condensation on the surface of chips and reduces influences of altered water concentration on its surface on the NIR spectra.

Ten representative wood chips were used as calibration standards for each hardwood species. It is in fact recommended for PCA discriminant analysis to have around 10 standards for each separate class. Unknown samples were then quantified based on these calibrations. The tested samples of wood chips labelled as „unknown“ were withdrawn from the same lot as standards. There were altogether 20 testing „unknown“samples for each hardwood species. NIR spectra of these chips were measured together with standards of the same hardwood species at approximately the same time period. By doing this, we achieved the conditions of standards to match those of unknown samples in order to eliminate any differences that might otherwise arise due to processes such as drying, decomposition or oxidation. These processes even if kept under control would cause small changes in chemical composition,; however even these small changes would have a great impact on the results interpretation, especially if chemical composition variations between hardwood species are so miniscule. If these processes were present during the storage, measurement of all hardwood chips at a short interval (within one or two days), and measurement of standards and unknown samples for each hardwood species at the same time eliminates any differences at which any undesirable processes might occur. Results suggest much higher selectivity and reproducibility of the results based on these particular measurements taken as compared to our previous results where this measure was not applied.

A disadvantage of this measure is of course the fact that once applied to online analysis, the recalibration with fresh standards might be troublesome due to the required frequency and precision. However, in order to achieve acceptable selectivity in online analysis technique development, it is important to consider this measure.

Instrument

Infrared spectrometer iS 10 by Nicolet was used for analysis. For NIR spectrometric measurements, Integrat IR extension by Pike Technologies was used.

Software

The NIR spectral data was collected using Omnic version 8.0 software by Thermo Scientific. The experiment setup was the following: Number of scans: 128; resolution: 4; data spacing 0.482 cm-1; final format: absorbance (Log (1/R)). The collected spectra were analyzed and evaluated using TQ Analyst version. 8.0 software by Thermo Scientific. The PCA method for discriminant analysis was incorporated in the TQ Analyst software.

RESULTS AND DISCUSSION

Infrared spectra of various species of hardwoods are almost identical and relevant differences except of different moisture content are invisible to the naked eye. The reason lies in very similar, almost identical, chemical composition, which includes celluloses, hemicelluloses, lignin, water and other, including extractives. In fact, large differences between the composition of hardwood species only lies in varying moisture content while variations between other chemical components are very small. However, even tiny differences in chemical composition of other components may be sufficient for successful distinction between hardwood species using PCA chemometrics discrimination.

In our previous studies (Russ et al. 2009), the success rate of positively identified samples was between 60-70 % with the exception of black locust, whose rate of positive identification was approaching 90 %. However, in our previous works, standards and unknown samples were not measured together but in a different time period. With the approach of measuring both standards and control samples, selectivity between hardwood species was greatly improved. Many PCA models were studied but only four are presented in this article.

1. In the first part, hardwood species were separated into two classes (class 1 and class 2) based on separation of hardwood species according to density, or hardness, respectively. The classes were chosen in the following way:
    Class 1: (high density wood): beech, turkey, oak, hornbeam, black locust
    Class 2: (low density wood): ash, birch, alder, aspen, poplar

2. In the second part, the hardwood species were separated in the same manner except for black locust, which was switched from class 1 to class 2. The PCA classification was changed in the following way:
    Class 1: (high density wood): beech, turkey, oak, hornbeam
    Class 2: (low density wood): ash, birch, alder, aspen, poplar, black locust

3. In the third part, the hardwood species were separated into the same two classes as in the first part, with the exception of ash, which was switched from class 1 to class 2. Classification of ash was on the borderline between high and low density hardwoods and therefore its role in selectivity improvement was investigated.
    Class 1: (high density wood): beech, turkey, oak, hornbeam, black locust, ash
    Class 2: (low density wood): birch, alder, aspen, poplar

4. In the fourth part, each hardwood species was a separate class by itself. In this case, ten separate classes were formed: beech, turkey, oak, hornbeam, ash, birch, alder, aspen, poplar and black locust.

1. PCA model

In the first PCA model, we separated the hardwood species into two groups with an emphasis to density based on our logical assumption. The following PCA diagram on Fig. 1 depicts separation of classes 1 and 2.

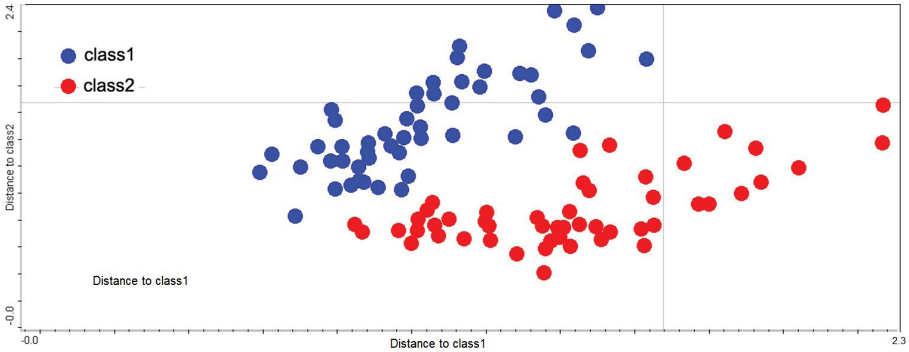

Fig. 1: PCA diagram of hardwood species of higher density (class 1, marked blue) versus hardwood species of lower density (class 2, marked red)

Fig. 1 depicts characteristic shift of hardwood species with higher density (blue dots) directing to the upper left quadrant. This shift was evident also in cases of higher density hardwoods in previous experimental studies. There is an obvious gap between points of the two classes which is an indication for distinctly higher selectivity.

Tab.1: Evaluation of unknown hardwood samples for the first PCA model

Wood species	Class 1	Class 2	Correct	Incorrect	Correct %
Black locust	19	1	19	1	95
Birch	1	19	19	1	95
Beech	17	3	17	3	85
Turkey oak	20	0	20	0	100
Oak	18	2	18	2	90
Hornbeam	19	1	19	1	95
Ash	6	14	14	6	70
Alder	0	20	20	0	100
Aspen	0	20	20	0	100
Poplar	0	20	20	0	100
Total	**100**	**100**	**187**	**13**	**96.5**

In Tab. 1 each of the first two columns contains counts of samples for each species evaluated as class 1 and class 2. Counts of correctly evaluated samples out of 20 are depicted in the third column. The last column represents correct sample classification in % (success rate of selectivity between the classes). The last row represents summary for all hardwood species.

The results in Tab. 1 show an exceptionally high selectivity based on PCA analysis illustrated on Fig. 1. In our previous studies, the success rate of positive identification between various hardwood species was ranging between 60-70 % with an expectation of black locust, which had a distinctly higher selectivity, likely due to the fact that black locust may have had a chemical component that the other hardwoods didn't have and was thus easily distinguished this way.

However, in the last study, we were able to obtain high increase of selectivity among all hardwood species as shown in Tab. 1. Only two out of ten hardwood species had lower success rate of identification than 90 %. Success rate above 90 % can be considered exceptional while only one or two out of twenty unknown samples are misidentified. One misidentified sample is not even significant and can be correlated with statistical error.

Our hypothesis of selectivity enhancement is based on different measurement approach already mentioned in the experimental part. Selecting chips as standards for calibration and unknown samples at the same time seems to be responsible for selectivity enhancement. Any minute changes resulting from decay, oxidation or dehydration that might arise due to additional storage, are eliminated. Even properly stored wood chips may eventually become discriminated with this sensitive method. It would in fact be possible that the differences in overall chemical composition that might arise due to additional storage of the same species may be greater than differences between different hardwood species. That would lead to inability to use this method for hardwood species discrimination. Despite very similar chemical composition of the studied hardwood species, the rate of successful identification of these species based on its density was 96.5 %.

2. PCA model

The second PCA model is in fact almost identical to the first one except for a switch of black locust from class 1 to class 2. This exchange of black locust was done on purpose. The purpose of this exchange was to study shift of overall success rate. In all of our previous studies, black locust has in all instances been separated into the upper-left quadrant. This quadrant of PCA model is where the higher density hardwood species tend to migrate. As mentioned, black locust exhibits the highest selectivity among all of the studied hardwood species and its movement to an unsuitable class should result in decrease of overall selectivity. As black locust became a component of the class 2, we expected a remarkable drop in overall selectivity, higher than exchange of any other components within both classes. By doing this, we could study the maximum possible induced error due to incorrect hardwood species identification that might be useful for further studies that would deal with the impact of improper hardwood classification. Fig. 2 depicts the situation with black locust switched from class 1 to class 2.

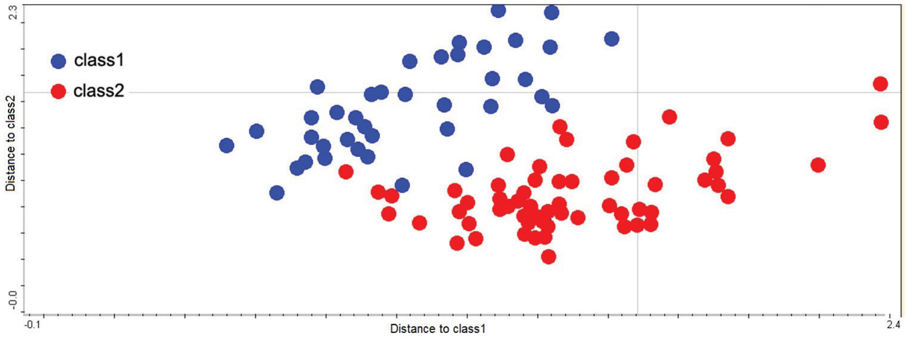

Fig. 2: PCA diagram of hardwood species of higher density (class 1, marked blue) versus hardwood species of lower density (class 2, marked red) with black locust switched position from class 1 to class 2

Fig. 2 clearly shows that PCA chemometrics analysis has changed and suffered somewhat due to black locust switch. The gap between the two clusters has become smaller and less defined. However, both classes still remain strong in their position and there seems to be no overlap between them.

Tab.2: Evaluation of unknown hardwood samples for the first PCA model

Hardwood species	**Class 1**	**Class 2**	**Correct**	**Incorrect**	**Correct %**
Black locust	4	16	16	4	80
Birch	8	12	12	8	60
Beech	8	12	8	12	40
Turkey oak	18	2	18	2	90
Oak	17	3	17	3	85
Hornbeam	18	2	18	2	90
Ash	1	19	19	1	95
Alder	1	19	19	1	95
Aspen	2	18	18	2	90
Poplar	1	19	19	1	95
Total	**78**	**122**	**164**	**36**	**82**

In Tab. 2 each of the first two columns contains counts of samples for each species evaluated as class 1 and class 2. Counts of correctly evaluated samples out of 20 are depicted in the third column. The last column represents correct sample classification in % (success rate of selectivity between the classes). The last row represents summary for all hardwood species.

The results clearly show that there is an overall decrease in selectivity from 96,5 % to 82 %. This 14.5 % drop is significant although it does not necessarily ruin the selectivity which is still sufficient for relatively positive identification. It is interesting to note, that selectivity is not related only to black locust but also to other hardwood species which have dropped in selectivity also. In fact, black locust rate of success has dropped from 95 % to 80 % which is only a difference of 15 %. The most suffered hardwood species turned out to be beech, whose success rate dropped from 85 % to 40 %, which accounts to 45 % difference. It is important to stress that switch of components between classes has an effect on the entire PCA model and thus all components are influenced by this because in each PCA model, they all influence one another. Due to a complicated mathematical calculation model which is generated by the software, it is not entirely possible to predict the outcome with relevant precision based on knowledge of other, independent PCA model behaviours.

3. PCA model

In order to discriminate hardwood species into two parts with the best possible separation we selected the first PCA model. In this model the ash, being on the borderline between low density and high density hardwood species, has been submitted to class 2. However, a more exact approach lets ash more qualified to the first class, thus belonging to the first class.

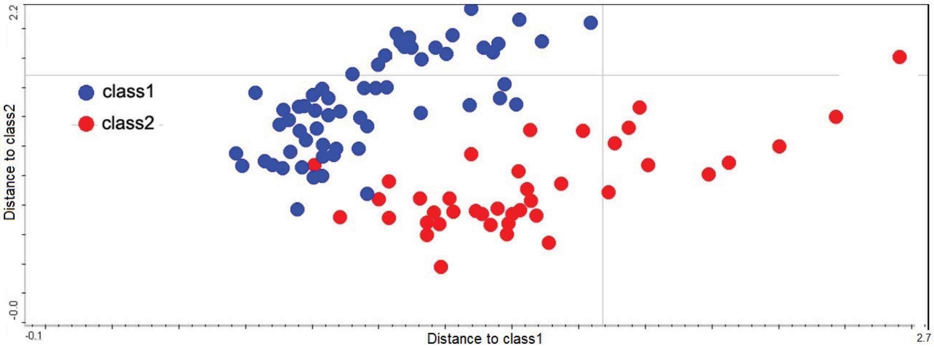

Fig. 3: Tensile strength and modulus of elasticity of early- and latewood versus the cambial age of annual rings

As Fig. 3 suggests, PCA model 1 is quite similar to PCA model 3 with both PCA diagrams being very similar. Unlike in PCA diagram 2, both PCA diagram 1 and 3 poses an obvious gap of free space between these two classes.

In Tab. 3 each of the first two columns contains counts of samples for each species evaluated as class 1 and class 2. Counts of correctly evaluated samples out of 20 are depicted in the third column. The last column represents correct sample classification in % (success rate of selectivity between the classes). The last row represents summary for all hardwood species.

As the results from Tab. 3 suggest, classification of ash as high density hardwood species instead of that of low density results in an identification success rate increase from 96.5 % to 98.5 %, which accounts for only 5 erroneous identifications out of 200 unknown sample trials.

Tab. 3: Evaluation of unknown hardwood samples for the third PCA model

Wood species	Class 1	Class 2	Correct	Incorrect	Correct %
Black locust	20	0	20	0	100
Birch	3	17	17	3	85
Beech	20	0	20	0	100
Turkey oak	20	0	20	0	100
Oak	20	0	20	0	100
Hornbeam	20	0	20	0	100
Ash	20	0	20	0	100
Alder	2	18	18	2	90
Aspen	0	20	20	0	100
Poplar	0	20	20	0	100
Total	78	122	164	5	98.5

In Tab. 3 each of the first two columns contains counts of samples for each species evaluated as class 1 and class 2. Counts of correctly evaluated samples out of 20 are depicted in the third column. The last column represents correct sample classification in % (success rate of selectivity between the classes). The last row represents summary for all hardwood species.

As the results from Tab. 3 suggest, classification of ash as high density hardwood species instead of that of low density results in an identification success rate increase from 96.5 % to 98.5 %, which accounts for only 5 erroneous identifications out of 200 unknown sample trials.

4. PCA model

While it is wise and a common approach to classify individual components into groups based on common properties within each group, we have studied a more complex PCA analysis where no such classes are created. Instead, we treated each hardwood species as a separate and independent class by itself in order to figure out whether it is still possible to distinguish between all of the individual hardwood species. Fig. 4 depicts PCA diagram of all ten hardwood species with each species classified as an individual class.

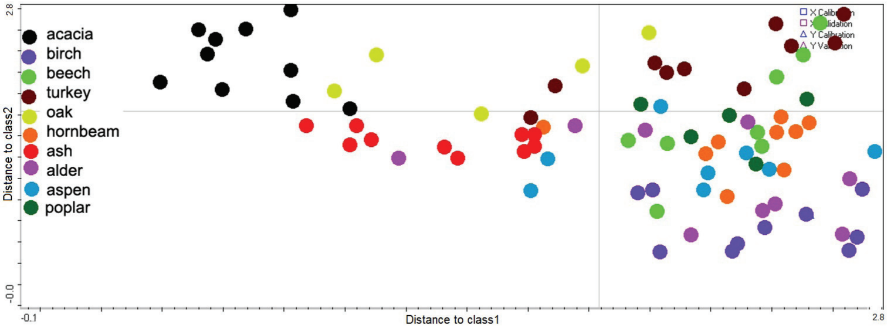

Fig. 4: PCA diagram of ten hardwood species selected as individual classes for each

Each class depicted in Fig. 4 represents a hardwood species on its own. In this case all ten hardwood species were included for PCA analysis. Usually, only about three or four classes are being used in order for PCA analysis to work properly. Adding more variables makes the mathematical model more complicated and it is more difficult for the mathematical model to find any trends that work sufficiently in distinguishing between each class. The PCA diagram suggests this complexity as most species inside the diagram overlap with neighbouring clusters. Again, it is apparent that black locust is well distinguished from the rest of the hardwood species. Other hardwood species are located in clusters that apparently follow the pattern of the previous PCA diagram. High density hardwood species have a tendency to shift into the upper left quadrant while lower density clusters descend into the lower right quadrant. This trend is witnessed in all PCA models we have evaluated during our research.

Tab. 4: Evaluation of unknown samples for hardwood species for the fourth PCA model

Wood species	**Black locust**	**Birch**	**Beech**	**Turkey oak**	**Oak**	**Hornbeam**	**Ash**	**Alder**	**Aspen**	**Poplar**
Black locust	**16**				3		1			
Birch		**13**	1			2		1	3	
Beech		1	**16**	2		1				
Turkey oak	1			**15**	2	2				
Oak					**17**		3			
Hornbeam				1		**19**				
Ash				1			**19**			
Alder		5					2	**11**		1
Aspen		3	1					2	**14**	
Poplar		1							1	**18**
Total	17	23	18	19	22	24	25	14	17	19
Error count 1	4	7	4	5	3	1	1	9	6	2
Error count 2	3	3	2	1	2	4	5	6	3	1
Errors	Total	%								
Error count 1	42	21								
Error count 2	30	15								

In Tab. 4 columns represent count of samples for each species. Rows represent counts of particular hardwood species. Based on complexity of the fourth PCA model, it was necessary to analyze the results in greater depth and consider different views of their interpretation. Besides counts of correctly determined samples, misclassified samples were all accounted for and each misclassified sample has been recorded as to which class it was identified as. This approach led to two different ways of success rate interpretation. A correct interpretation is based on the number of errors during each series of samples classification. A correct success rate classification lists how many times a sample was identified correctly out of total. This approach was used in the first through third PCA model. In this case, the total success rate of correct identification was 79 % (21 % of unknown samples were identified incorrectly).

If we are interested in total counts of hardwood samples where some species samples are misidentified but both ways around so that eventually cancel each other out. Such interpretation leads to decrease of error by 6 % to 15 % and makes the success rate as high as 85 %. Continuous online analysis of wood chips such as those on transport belts relates to this type of result interpretation well since samples are randomly picked out and only the total counts for each hardwood species is taken into consideration. Success rate of 85 % for a PCA diagram of 10 variables is considered exceptional.

CONLUSIONS

The results of our study suggest that it is both possible to successfully distinguish hardwood species based on their density differences as well as minute chemical composition variations among individual hardwood species. PCA diagrams as well as other presented results prove good selectivity among both classes of selected hardwood chips and also as many as ten individual hardwood species between each other.

Our previous studies showed that the selectivity of black locust was the highest among all hardwood species we dealt with. It was the only hardwood species which had sufficiently high selectivity. However, in this study all hardwood species had much higher selectivity and the standout of black locust was no longer obvious.

We explain our profound gain in selectivity with a different approach in sample preparation and analysis. All hardwood standards and samples were selected from a single lot and were processed and measured at the same time, minimizing any additional unwanted chemical composition changes.

Separation of hardwood species based on their density and classifying them according to the first PCA model led to classification success rate of 96.5 %. This success rate was much higher than we previously expected and only proved that NIR spectroscopy combined with chemometrics evaluation was one of the most suitable methods for the specific purpose. Based on the high success rate we expect this method to be used for online evaluation of certain hardwood species in form of chips before pulping processes.

Separation of each individual hardwood species into independent components led to lower success rate of 79 %. However, considering the challenge in software calculation as well as the amount of hardwood species in the diagram, this rate cannot be considered low. Still, due to switched misidentification on the long run, only the total chip counts matter, not every single independent chip identification. Considering this, the selectivity success rises to 85 %. This calculation would be used for the selectivity success rate that would be applicable for online analysis.

ACKNOWLEGMENT

This work was supported by Slovak Research and Development Agency under contract No. APVV-0338-07.

REFERENCES

1. Michell, A.J., Schimleck, L.R, 1996: NIR spectroscopy of woods from Eucalyptus globules. Appita Journal 49: 23-26.
2. Russ, A., Fišerová, M., Gigac, J., 2009: Preliminary study of wood species identification by NIR spectroscopy. Wood Research 45(4): 23-32.
3. Russ, A., Fišerová, M., Gigac, J., 2010: Ability of NIR spectroscopy to identify hardwood species. Papír a celulóza 65 (12): 364-367.
4. Schimleck, L.R., Wright, P.LJ., Michell, A.J., Wallis, A.F.A., 1997: Near infrared spectra and chemical compositions of E-globulus and E-nitens plantation woods. Appita Journal 50: 40-46.
5. Schimleck, L.R., Evans, R., Ilic, J., 2001: Application of near infrared spectroscopy to diverse range of species demonstrating wide density and stiffness variation. IAWA Journal 22: 415-429.
6. Tsuchikawa, S., 2007: A review of recent near infrared research for wood and paper. Applied Spectroscopy Review 42 (2): 43-71.

Albert Russ
Pulp and Paper Research Institute
Lamačská Cesta 3
841 04 Bratislava
Slovak Republic
Corresponding author: russ@vupc.sk

Mária Fišerová
Pulp and Paper Research Institute
Lamačská Cesta 3
841 04 Bratislava
Slovak Republic
E-mail: fiserova@vupc.sk