

Functional nonparametric classification of wood species from thermal data

Javier Tarrío-Saavedra · Salvador Naya ·
Mario Francisco-Fernández · Jorge López-Beceiro ·
Ramón Artiaga

ISBC XVI Conference Special Issue
© Akadémiai Kiadó, Budapest, Hungary 2010

Abstract In this study, thermogravimetric (TG) and differential scanning calorimetry (DSC) curves, obtained by means of a simultaneous TG/DSC analyzer, and statistical functional nonparametric methods are used to classify different wood species. The temperature ranges, where the highest probability of correct classification is reached, are also computed. As each observation is a curve, a nonparametric functional discriminant technique based on the Bayes rule and the Nadaraya–Watson regression estimator is used. It assigns a future observation to the highest probability predefined class (supervised classification). The smoothing parameter needed in this nonparametric method is selected according to the cross-validation technique. The method proposed is applied to a sample of 49 wood items (7 per wood class) and also to classify between hardwoods and softwoods. In all the cases, the samples have been successfully classified, obtaining better results with the TG curves. The results are compared with those obtained with

other nonparametric methods based on boosting algorithm. A discussion about the relation of the obtained results with the referenced wood component degradation temperature ranks is presented.

Keywords Wood · Nonparametric classification · Functional data analysis · Thermal analysis

Introduction

The identification of wood is one of the most difficult tasks to perform related with the technology of this material, due to the wide variety of species and anatomical heterogeneity of its elements. Wood identification can often be made on the basis of readily visible characteristics such as color, odor, density, presence of pitch, or grain pattern. This analysis is typical in the furniture industries and the wood panel production. Often, the performed analysis has a non-uniform accuracy because of the operator. To achieve a correct classification is essential to use microscopy techniques, physical hardness tests, and chemical analysis [1–3]. Therefore, the implementation of quantitative models and automatic recognition methods of wood samples are justified and can be immediately useful. While there are various computational procedures to evaluate and rate the quality of a timber by inspecting its defects with the help of image processing techniques and spectral analysis [1, 4–8], these are not so generally used for species identification, despite several studies addressing this problem exist [2, 9–13].

A first step in a classification problem is to choose a discriminant feature from which it will be possible to classify. In the case of wood species classification, this discriminant feature could be the output of an experimental technique that really differentiates between them. In the

J. Tarrío-Saavedra (✉) · M. Francisco-Fernández
Departamento de Matemáticas. Facultad de Informática,
Universidade da Coruña, Corunna, Spain
e-mail: jtarrío@udc.es

M. Francisco-Fernández
e-mail: mariofr@udc.es

S. Naya
Departamento de Matemáticas. Escuela Politécnica Superior,
Universidade da Coruña, Corunna, Spain
e-mail: salva@udc.es

J. López-Beceiro · R. Artiaga
Departamento de Ingeniería Industrial II. Escuela Politécnica
Superior, Universidade da Coruña, Corunna, Spain
e-mail: jlopezb@udc.es

R. Artiaga
e-mail: rartiaga@udc.es

literature, wood samples are mainly classified based on the results of two techniques: image- and spectrum-based processing systems. A method of classification of 20 types of tropical timber from image processing, using extracting textural wood features, has been successfully tested in [9]. Those authors obtained a good classification proportion of 95%. On the other hand, in [2], the Fourier Transform Raman (FTR) spectroscopy and neural network technology have been coupled for spectral feature extraction and non-supervised classification. This represents the first time that both methodologies are combined. Later, neural networks and the FTR spectra for hardwoods and softwoods to differentiate temperate woods from tropical woods were also used [13]. Genetic algorithms and principal components analysis were used to classify 98 Raman spectra of temperate softwoods, hardwoods, and Brazilian and Honduran tropical woods [10]. Recently, in [12], an automatic wood type classification system based on the analysis of the fluorescence spectra, using nearest neighbor classifiers, linear and quadratic classifiers, and support vectors machines (SVMs) has been designed. However, it seems that the possibility of using thermal analysis as a source of data for statistical classification of wood species has not been sufficiently studied yet. In this article, the thermograms obtained by thermogravimetric analysis (TG) and differential scanning calorimetry (DSC) are used as a discriminant characteristic. These curves can be processed in a relatively simple way with functional analysis [14, 15, 16] and, as shown below, the shape of the TG curves is directly related to the wood composition. Therefore, TG analysis becomes an interesting option to discriminate between classes of timber.

In general, wood is defined as the set of xylem tissues forming the trunk, roots, and branches of woody plants, excluding the bark. The tubular cells size, shape and distribution, along with other anatomical elements such as wood radii, the presence of resin canals or vessels, pores, etc., in addition to the variable proportion of its chemical components, define the different wood species and their properties [1, 3, 13, 17–23]. Also, the different wood types can be generally divided in two broad categories: softwoods or conifers (gymnosperms) and hardwoods (dicot angiosperms) which can be subdivided into boreal, austral, and tropical hardwood types [1, 3]. Is it possible to observe these differences among species in the shape of the TG curves in a pyrolysis test? According to existing studies, the answer is yes [13, 17–23].

The wood degradation in an inert atmosphere is dominated by the degradation behavior of its three main components [17]. These are cellulose, lignin, and hemicellulose [13, 17–24]. Cellulose represents about 40 and 60% in the overall weight of dry wood (it accounts for 23–33% of the mass of softwoods), 23–33% of lignin in softwoods

(16–25% in hardwoods), and 25–35% of hemicellulose (more in hardwoods than in softwoods) [3, 19, 25]. The three components decompose in temperatures ranging between 240–350 °C, 280–500 °C, and 200–260 °C, respectively [17, 19–23, 25]. As was reported in [13, 18], the TG curve describing the pyrolysis of wood nearly coincides with the sum of the degradation of its constituents. In many cases, no significant interaction between them has been concluded [21]. The proportion of each wood component varies depending on the species, to a greater or lesser extent [19, 20, 22, 23]. Therefore, it is expected that the TG curves are different depending on the type of wood to which they belong. While the effect of the wood structure appears to exist [22], this is much lower than that of the components [13, 18, 21]. Furthermore, differences in the pyrolysis of lignin and hemicellulose depending on whether these come from softwood or hardwood, or even of different species, were observed [20, 23]. These results suggest the use of discriminant characteristic TG curves.

Accordingly, the objectives of this study are

1. Evaluating the potential of functional nonparametric methods of discriminant analysis for the classification of hardwoods and softwoods and then for the classification of European oak, European chestnut, eucalyptus, scots, and insignis pine on the basis of TG and DSC data.
2. Comparing the accuracy performance of TG or DSC curve classification to discriminate between wood species or between major groups. The supervised kernel nonparametric classification and kernel nonparametric classification using the k -nearest method to select the bandwidth h are used in [14]. In addition, two methods based on the boosting algorithm are also used to complete the study: using principal components analysis (PCA) and by the representation of functional data on a b -spline basis [15, 16, 26].
3. Finding the temperature range in TG and DSC curves where the highest probability of correct classification is reached.
4. Relating the results of classification analysis in each interval with the referenced cellulose, lignin, and hemicellulose degradation temperature ranks in a nitrogen atmosphere.

Experimental

Materials

Tests for five different hardwoods (European beech or *Fagus sylvatica*, European oak or *Quercus robur*, chestnut or *Castanea sativa*, *Eucalyptus globulus*, and jatobá or

Hymenaea courbaril) and two softwoods (scots pine—*Pinus silvestris* and insignis pine—*Pinus radiata*) are carried out. Seven samples per each one of the above mentioned species, obtained from wood of different trees are tested. The aim of this sampling process is to obtain a compromise between capturing the existing variability and minimizing the time of experimentation. The samples are not dried to avoid disturb as much as possible their structure and composition, and to test the automatic classification method with a minimal sample preparation, under the worse conditions.

Measurement methods

The test is performed on a SDT 2960 TA Instruments thermo balance. This apparatus provides both TG and DSC curves used in the classification analysis. A heating ramp of 20 °C min⁻¹ is applied in the range from 20 to 600 °C at a rate of 50 mL min⁻¹ of N₂ [27]. The nitrogen is purged for 10 min, before starting the heating program for establishing an inert environment. The used heating rate is chosen to obtain a proper balance between time test and resolution [27]. It aims to assess the discriminatory power of the resulting curves, using the minimum experimental time. The sample mass chosen is between 6 and 8 mg. Alumina crucibles are used. In particular, TG and DSC measurements are affected by some experimental parameters such as heating rate, amount of mass, type of atmosphere, or sample geometry [27]. Therefore, all these parameters are remained constant to obtain a better classification.

Classification techniques

Nonparametric functional techniques based on kernel methods [14] and two nonparametric methods based in the boosting algorithm are applied to construct a classification rule to discriminate between hardwoods and softwoods, and between the different species: European beech, European oak, European chestnut, eucalyptus, jatobá, scot, and insignis pine, based on a sample of 49 TG and DSC curves. A DSC or TG curve is classified as belonging to the specie or the group to which the highest posterior probability is obtained.

The functional Nadaraya–Watson kernel nonparametric method (K-NPFDA), shown in (1), is applied. Given a new TG or DSC curve, $x = x(t)$, obtained from a material to classify, the estimator of the posterior probability of belonging to a class g , with $g \in \{0, 1, \dots, G\}$, is given by:

$$\hat{r}_h^{(g)}(x) = \frac{\sum_{i=1}^n I_{\{Y_i=g\}} K\left(\frac{\|x-X_i\|}{h}\right)}{\sum_{i=1}^n K\left(\frac{\|x-X_i\|}{h}\right)}, \tag{1}$$

where the observed TG/DSC curves, $X_i = X_i(t)$, are a sample of explanatory variables, while the response sample

consists of the observations Y_i of a discrete random variable taking values in the set $\{0, 1, \dots, G\}$, the different classes. The parameter h is the bandwidth or smoothing parameter and $\|\cdot\|$ denotes the following distance between curves:

$$d(X_i, X_j) = \int_a^b (X_i(t) - X_j(t))^2 dt, \tag{2}$$

where $[a, b]$ is one of the 1,280 temperature intervals studied.

After a careful examination of the TG/DSC curves, further processing of the data has been found useful for standardizing the curves [28]. Denoting by $f(x)$ a curve, a linear transformation, $\tilde{f}(x) = \alpha f(x) + \beta$ with

$$\alpha = \frac{\sqrt{b-a}}{\sqrt{\int_a^b \left(f(t) - \frac{1}{b-a} \int_a^b f(s) ds\right)^2 dt}}$$

and

$$\beta = 1 - \frac{\int_a^b f(t) dt}{\sqrt{b-a} \sqrt{\int_a^b \left(f(t) - \frac{1}{b-a} \int_a^b f(s) ds\right)^2 dt}},$$

is done to achieve

$$\frac{1}{b-a} \int_a^b \tilde{f}(t) dt = 0$$

and

$$\frac{1}{b-a} \int_a^b \left(\tilde{f}(t) - \frac{1}{b-a} \int_a^b \tilde{f}(s) ds \right)^2 dt = 1$$

This transformation should act on the mean and variance to improve the discriminant power of the curves.

In our research, the Gaussian kernel, K , is used. On the other hand, the smoothing parameter, h , is chosen as the value that minimizes the probability of misclassifying a future observation and it is selected according to the cross-validation method [29]. This method consists in minimizing the cross-validation function:

$$CV(h) = n^{-1} \sum_{i=1}^n I_{\{Y_i \neq d_h^{-i}(X_i)\}},$$

where d_h^{-i} is the classification rule built up without the i -th observation:

$$d_h(x) = \operatorname{argmax}_{0 \leq j \leq G} \left\{ \hat{r}_h^{(j)} \right\}.$$

It can be useful and efficient to replace the h parameter, a real number, by an integer parameter k from a finite subset.

A way to achieve this is to consider a k Nearest Neighbors (kNN) version of the kernel estimator [14]. In this study, it is named KNN-NPFDA method. The number of neighbors and the bandwidth is selected using the cross-validation method.

When the kNN estimator is used, the parameter h is replaced by h_k , which is the bandwidth allowing us to take into account k terms in the weighted average [14]:

$$\hat{p}_k^{(g)}(x) = \frac{\sum_{\{i:Y_i=g\}}^n K(h_k^{-1} \cdot d(x, X_i))}{\sum_{i=1}^n K(h_k^{-1} \cdot d(x, X_i))},$$

where h_k is a bandwidth such that

$$\#\{i : d(x, X_i) < h_k\} = k,$$

with $\#$ the cardinal of the set.

Two additional nonparametric methods based on the boosting algorithm, the B method and the B-PCA method, are implemented to be compared with the kernel methods. They are specially designed to perform a nonparametric supervised classification for functional data. The boosting algorithm used is the Adaboost algorithm for classification [26].

In the B method, the boosting algorithm estimates the optimal number of basis and the optimum depth of the tree partition using Functional Data Object for obtaining the best possible estimation.

In the B-PCA method, the Adaboost algorithm is applied to a set of data using Principal Component Analysis. The optimal number of Principal Components and the optimum depth for one or more classifiers are estimated.

The free statistical software R [30] is employed to implement the nonparametric functional methods used in this article. Mainly, the R packages `fda` and `fda.usc` are used to perform the classification applying nonparametric functional analysis. Adaboost algorithm is a modified algorithm of the function `adabag.M1` of `adabag` R package that fits adaboost algorithm with classification trees as individual classifiers.

Results and discussion

In this section, the methods previously presented are applied to the TG and DSC curves to classify between different species and main groups. First, a descriptive analysis of the data is shown. It is important to note that each method is validated through cross-validation, which is the technique widely used for the validation of an empirical model. It works by leaving out one TG/DSC curve; then a model is trained with the remaining thermograms and, finally, the developed model is used for the classification of the left out TG/DSC curve. This is repeated until all the curves have been left out once. As the data set available contain 49 samples, 48 samples are used for training and 1

sample for testing. This process is repeated 49 times, and the percentage (measured as per one) of correct classification are calculated.

Descriptive analysis of the TG curves

Figure 1 shows the 49 TG curves obtained (7 per class). As can be observed, each curve represents a functional data. A particular wood specie is highlighted in each panel. The last panel (in row 4, column 2) shows all the TG curves. Different trends are observed for almost all types of wood. Even so, the variability in each class prevents from discerning clearly in some intervals. Apart from this, some species such as oak and beech tend to overlap. This is correlated to their similar densities, hardness and mechanical properties. In addition, they belong to the broader group of boreal hardwood [1].

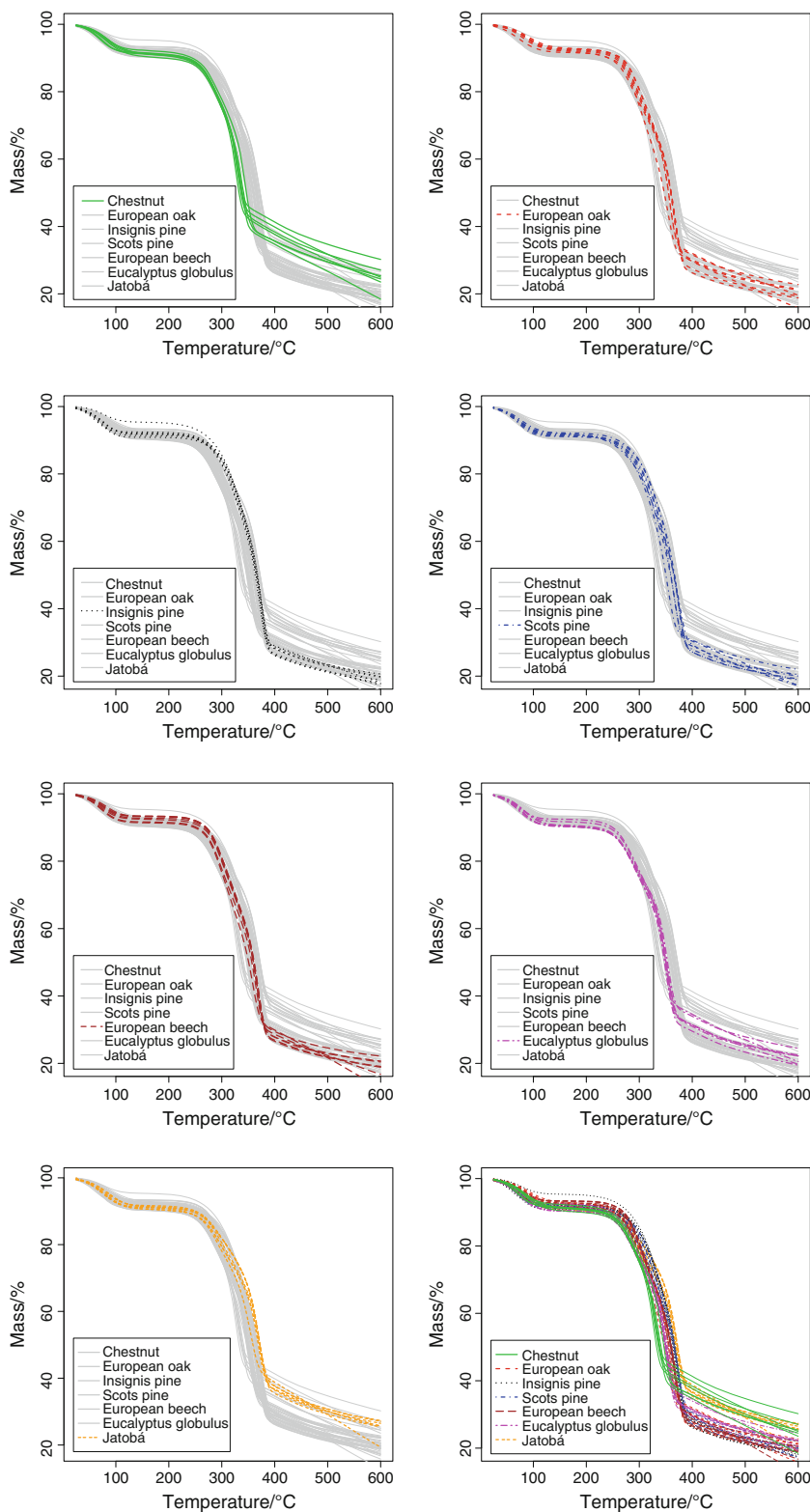
In Fig. 2a, the means of the TG curves for each class of wood, as defined by Fraiman and Muniz [31], are plotted. They can report on the possible degree of overlapping among the degradation trends for the different species and in what intervals this happens. In fact, it appears that differences between species are starting to take place from 200 °C onwards, coinciding with the beginning of the hemicellulose degradation [22, 25]. These differences become maxima in the range of temperatures where the cited maximum decomposition rate of the cellulose and lignin occurs [13].

As the storage period is long enough (over a year) and the storage conditions of all wood samples are the same, through the TG curves is possible to measure the water absorption capacity of each timber. In fact, it is observed that the height of the first step is slightly different for some species, being able to build two groups: oak, beech, insignis pine and, on the other hand, chestnut, eucalyptus, and jatobá. It can be observed that the existing residue in the range of 400–600 °C is different depending on the species (chestnut and jatobá > eucalyptus > oak, scots pine, beech and insignis pine).

Figure 3a shows the variability in each class. By working with functional data, the variance of the TG curves for a class is not a value, but a curve. The greatest variability occurs in the range of temperatures where the maximum decomposition rate of the cellulose and lignin is produced, according to Yang et al [13]. Therefore, the use of classification methods should work worse at these temperatures (320–370 °C, see Fig. 3a), but this depends on the magnitude of the variability among species in this interval, which is also higher (see mean differences in Fig. 2a). This variability is expected given the heterogeneity of wood [1, 3].

Figure 4 shows the DSC curves obtained for every sample tested, in the same form as in Fig. 1. There is a

Fig. 1 Original TG curves (7 per class), where each particular wood specie is *highlighted* in the corresponding *panel*. In the last *panel* (row 4, column 2) all the *curves* are presented



wide variability in each class, higher than in the TG curves. Therefore, it is expected that the DSC curves discriminate worse than TG between classes of wood. According to the

literature, the endo and exo DSC events of hemicellulose, cellulose, and lignin overlap in the 220–520 °C range [13, 22, 25]. The differences among species can be set

Fig. 2 **a** Functional means of the original TG curves for each specie (*left panel*). **b** Functional means of the location-scale transformed TG curves (*right panel*)

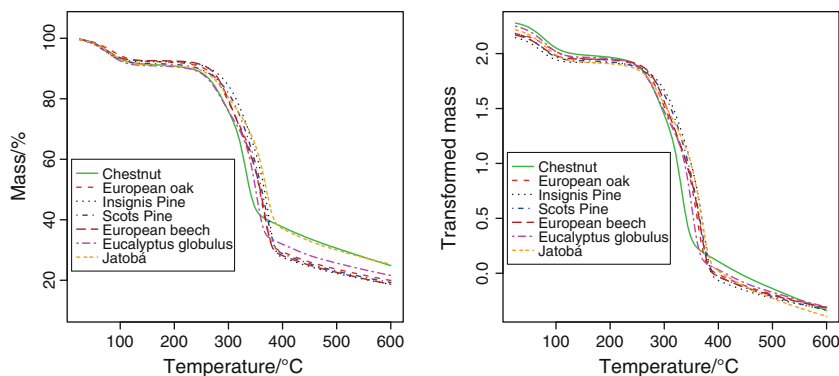
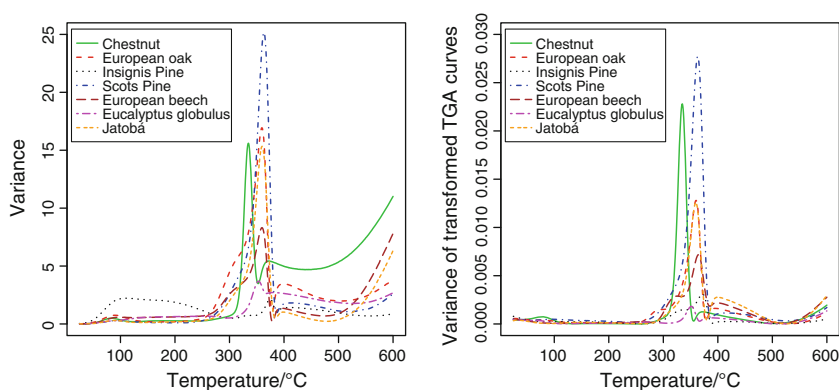


Fig. 3 **a** Functional variances of the original TG curves for each specie (*left panel*). **b** Functional variances of the location-scale transformed TG curves (*right panel*)



according to the displacement and magnitude of these three previous peaks (see Fig. 5), which may be due to different weight percentage of hemicellulose, cellulose and lignin, interactions, or differences in the structure [19, 20, 22, 23]. There are also differences in the peak corresponding to the water loss (see Fig. 5). The classification methods implemented may determine whether these differences are reliable.

By implementing a location-scale transformation, a reduction of the variance in each class is sought. Figures 6 and 7 show the TG and DSC data after this processing step. A reduction in the dispersion of the curves in each class is observed (Fig. 3). There is also an increasing distance between curves from different classes in certain temperature ranges.

Result of the data transformation

We apply nonparametric kernel methods for functional data (K-NPFDA and KNN-NPFDA) and methods based on boosting algorithm (B and B-PCA). It is observed that the transformation of the data [28] significantly helps to distinguish among the different groups (hardwoods and softwoods) and also among species. For example, if the TG

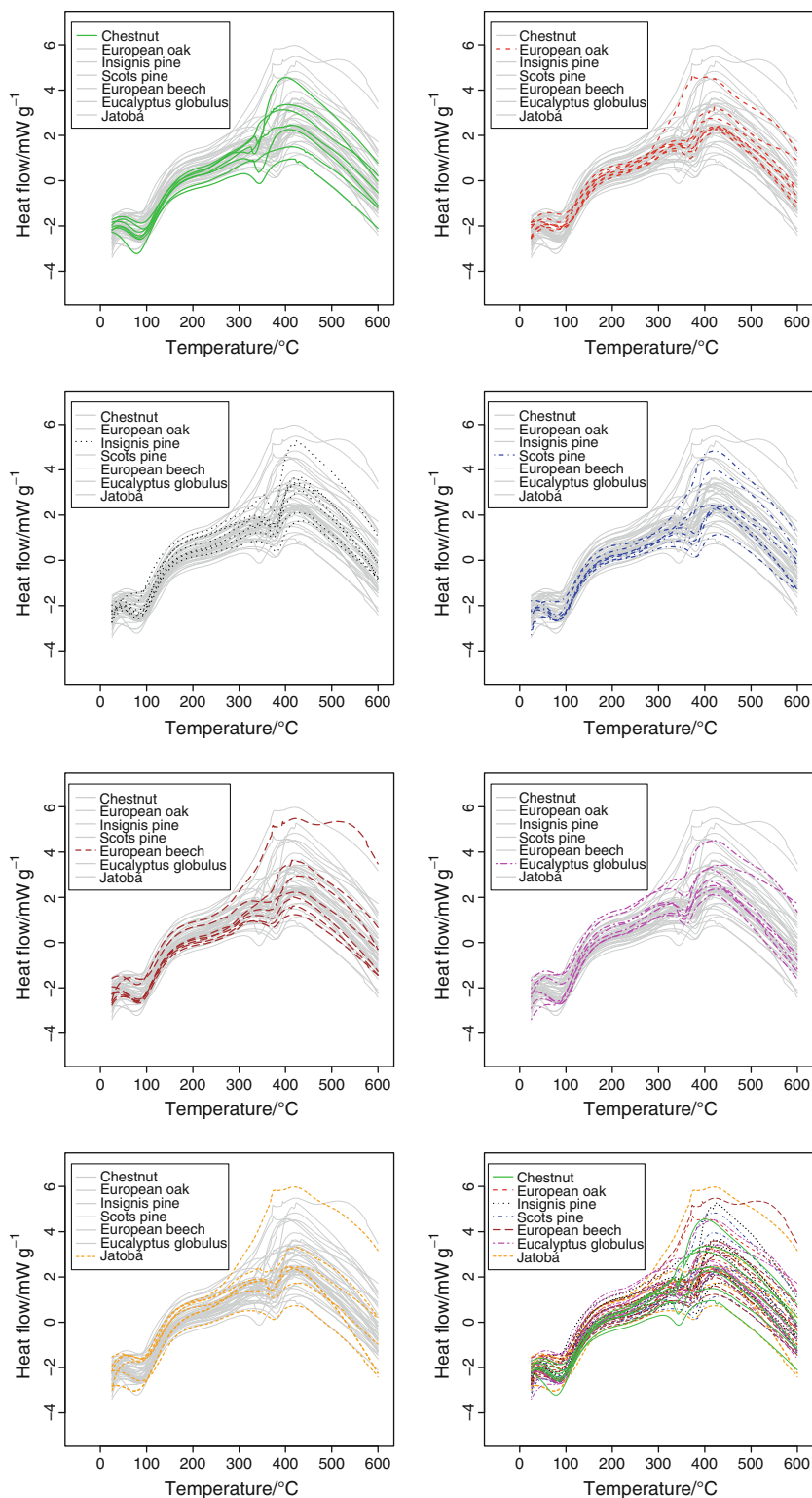
curves are analyzed by the K-NPFDA method, classifying samples from seven kinds of wood, then a probability of correct classification equal to 0.79 is obtained in the best of possible intervals (180–330 °C); on the other hand, if we use the transformed data, then a probability of 0.88 is obtained in the range 192.5–292.5 °C. This is repeated for all models and data analysis; therefore, the results using the transformed data are shown, see Table 1.

TG curve classification

In Table 2 the probabilities of correct classification and the temperature ranges for which they are maxima are shown. They are computed in two settings, classifying among the seven different species and in the more general case of classifying into three different groups. This is the result of evaluating the probabilities at 1,280 intervals of eight different sizes, from 50 to 400 °C. The four methods of classification described above are calculated.

The results in Table 2 can be grouped in two blocks, those techniques based on a kernel and the Nadaraya–Watson estimator and those based on boosting algorithm. In the case of classification in three groups or general classes

Fig. 4 Original DSC curves (7 per class), where each particular wood specie is highlighted in the corresponding *panel*. In the last *panel* (row 4, column 2) all the curves are presented



of wood, the K-NPFDA (with $0.0005 < h < 0.001$) and KNN-NPFDA techniques, with an optimal number of neighbors equal to 1, are the best methods with an estimated probability of success of almost equal to 1 (0.94). It is

interesting to note that the four methods practically coincide in the optimal range (192.5–292.5 °C and 210–310 °C), which in turn coincides with the region of hemicellulose degradation reported by several authors [22, 25]. The

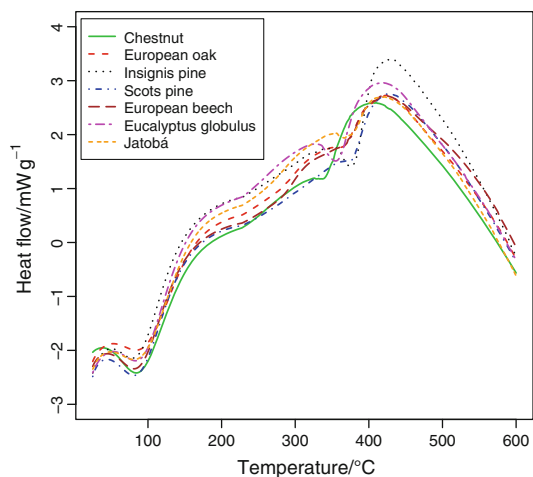


Fig. 5 Functional means of DSC curves

hardwoods tend to have a higher content of hemicellulose [3, 19, 25] and their weight proportion influences the total degradation of wood. Classification between hardwoods and conifers has been successfully achieved. Also, the correct classification among austral, boreal and tropical hardwoods is obtained. In fact, there is very small confusion between the groups. In Table 3, the classification matrices in the optimal intervals presented in Table 2 using 3 different classes (boreal hardwoods, softwoods and other hardwoods) are shown. It can be observed that using the K-NPFDA method the probability of classifying a boreal as a tropical or austral hardwood is only 0.05 and the probabilities of correct classification in each group are very high in this case (0.95 for boreal hardwoods, 0.86 for softwoods and 1 for other hardwoods). This result is similar to that obtained by other techniques using image- and spectrum-based processing systems [2, 9–13].

The overall probabilities of correct classification when one wants to discriminate among the seven existing species of wood is also very high. Especially interesting are the results obtained using the methods K-NPFDA and B (7 elements in the basis and depth of tree equal to 3). In the first case, a probability of correct classification of 0.88 for the interval 192.5–292.5 °C is obtained (see Table 2). This could be due to the different hemicellulose content and differences in hemicellulose degradation depending on the species, but we must do more experiments to prove it. The optimal interval (217.5–417.5 °C) obtained by the method B includes the degradation processes of the hemicellulose, cellulose and lignin getting a slightly higher probability of correct classification. In fact, according to Müller [20], the differences in wood species are mainly due to the different thermochemical behavior of lignin degradation and that of the first step of the hemicellulose degradation.

Table 4 shows the classification matrices in the optimal intervals presented in Table 2 using seven different classes. It can be observed that the results obtained by Methods B and K-NPFDA are complementary. The probabilities of correct classification of each kind of wood are relatively high in both cases. The K-NPFDA method only fails to discriminate the scots pine class ($P = 0.58$). Instead, a probability of correct classification equal to 1 of the scots pine class is obtained by the B method. Among other things, it can be due to the much larger optimal temperature range, encompassing much of the lignin degradation which provides more information to differentiate the species. On the other hand, the K-NPFDA method classifies slightly better oak and jatobá woods than the B method (see Table 4).

In general, it appears that the cited range of decomposition of the hemicellulose and, to a lesser extent, of lignin and cellulose is the temperature range where more differences between species exist.

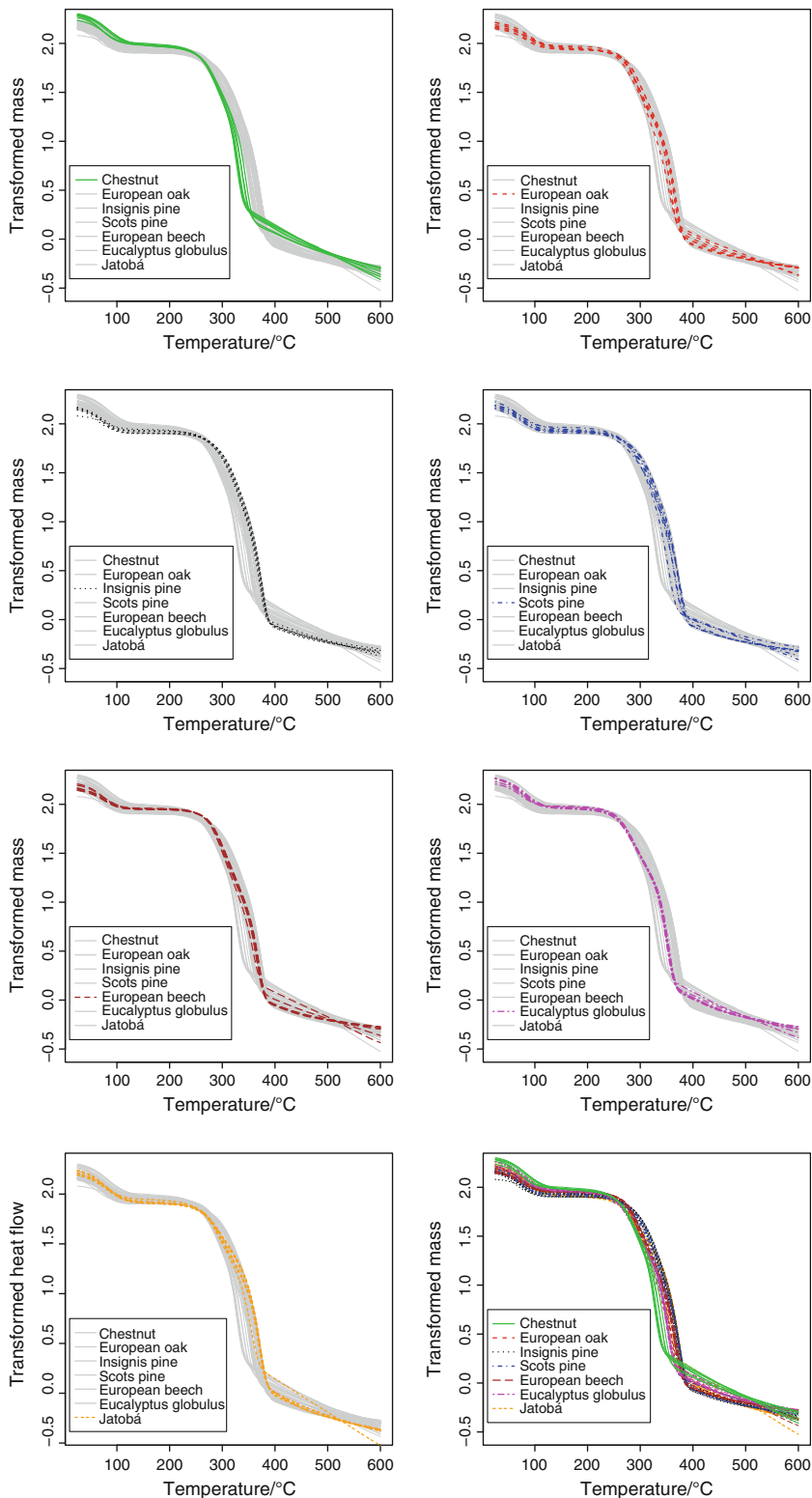
Apart from using leave-one-out cross-validation, the prediction power of the K-NPFDA method is measured. For this, the whole set of 49 curves is divided into two groups: a training sample of 42 curves, and a test sample of seven curves (one for each class of timber). Our aim is try to classify correctly the test sample using the training sample. This problem is more common in industry. Table 5 shows the classification matrix obtained when classifying among the seven considered species. In this table, the results in the temperature interval of 207–307 °C are shown. This is the optimal interval using cross-validation with the training sample. It can be observed that the 100% of the test sample is successfully classified using the K-NPFDA method. Same results for the case of classifying among the three main groups are shown in Table 6 (in the interval 192.5–292.5 °C), obtaining the same success.

In conclusion, it is proved that using TG curves as discriminant characteristic is possible to classify different species of wood.

DSC curves classification

Tables 7, 8, and 9 show similar results to those presented in Tables 2, 3, and 4, respectively, but using the DSC curves obtained by a simultaneous SDT. It can be observed that DSC curves have less discriminating power than TG curves. Nevertheless, very good results are obtained when we try to distinguish among boreal hardwoods, softwoods, and tropical or austral hardwoods (Tables 7 and 8). The kernel nonparametric functional methods (K-NPFDA and KNN-NPFDA) work better than the others based on the boosting algorithm. A good classification probability equal to 0.80 is obtained in the temperature range of 322.5–485 °C. This

Fig. 6 Location-scale transformed TG curves

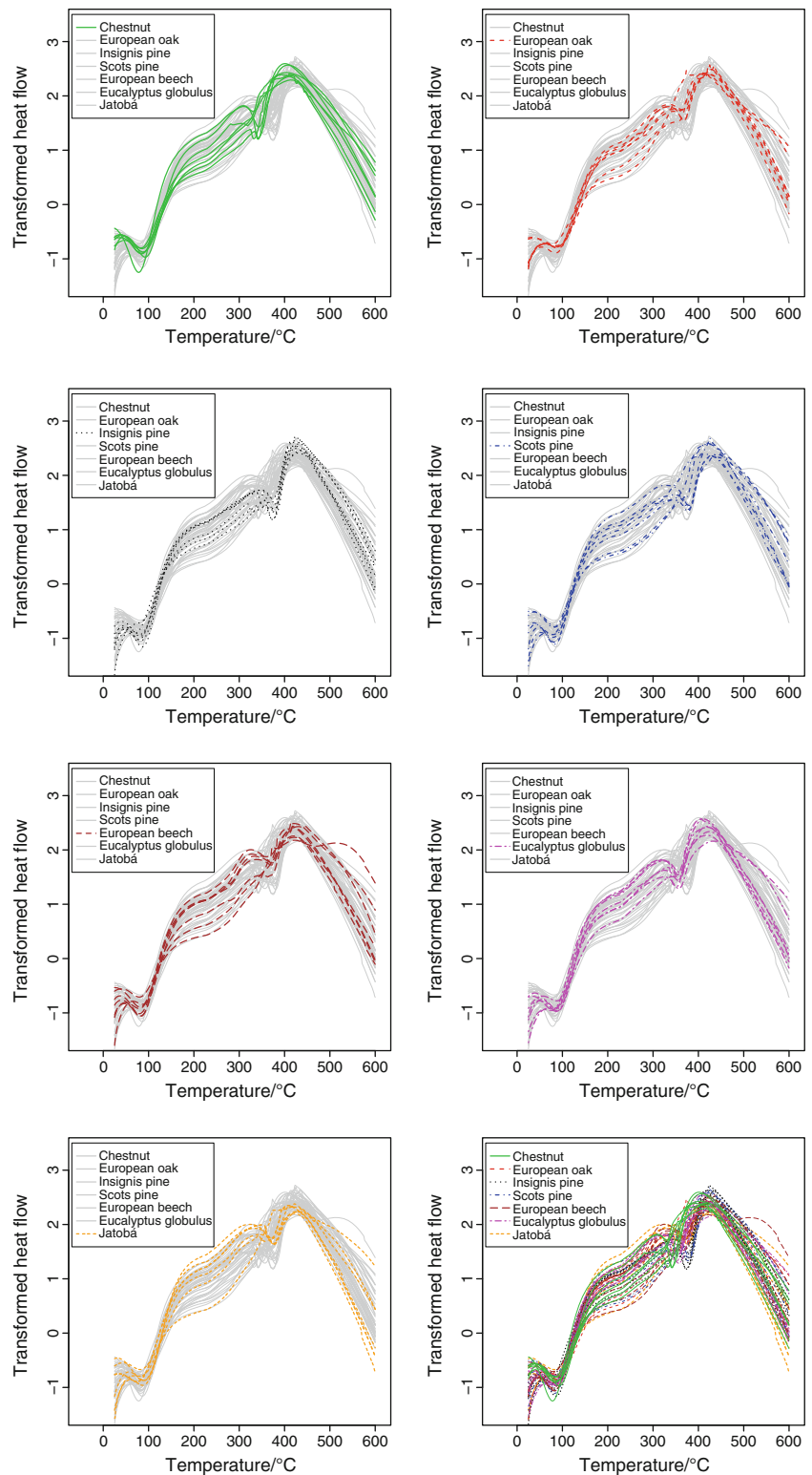


interval corresponds to the region of maximum degradation rate of cellulose and lignin, reported by several authors [13, 22, 25]. Therefore, the DSC curves have a higher classification power in this area; there are more differences between

species. These differences may be due to the nature and weight percentage of lignin [11].

When we want to classify among the seven species, the DSC results are worse than the TG ones. The best

Fig. 7 Location-scale transformed DSC curves



methods are again K-NPFDA and KNN-NPFDA with a maximum probability of correct classification equal to 0.60 for the interval 322.5–347.5 °C. Thus, using the K-NPFDA method, good classification results are only obtained in the case of eucalyptus, chestnut and jatobá

(Table 9) but the two kinds of pine are often misclassified and the beech curves are often classified as oak curves ($P = 0.29$). In fact, both pairs of species have very similar mechanical properties, hardness and density [1].

Table 1 Correct classification probabilities using original and transformed data, with 3 or 7 classes

Data	Original data		Transformed data	
	Number of classes	Corr. class. prob.	Number of classes	Corr. class. prob.
TG	3	0.94	3	0.94
TG	7	0.79	7	0.88
DSC	3	0.42	3	0.79
DSC	7	0.24	7	0.57

Table 2 Correct classification probabilities and optimal intervals obtained by each classification method

Methods	3 Groups classification		7 Groups classification	
	Optimal interval/°C	Corr. class. prob.	Optimal interval/°C	Corr. class. prob.
K-NPFDA	192.5–292.5	0.94	192.5–292.5	0.88
KNN-NPFDA	192.5–292.5	0.94	182.5–282.5	0.88
B	210.0–310.0	0.90	217.5–417.5	0.90
B-PCA	210.0–310.0	0.90	195.0–345.0	0.83

The TG data were tested with 3 (boreal hardwoods, softwoods, other hardwoods) and seven classes

Table 3 Classification matrices using 3 different classes (boreal hardwoods, softwoods, and other hard woods) obtained by different non-parametric classification methods, using TG data

Methods	Estimated	Theoretical		
		Boreal hardwoods	Softwoods	Other hardwoods
K-NPFDA	Boreal hardwoods	0.95	0.07	0.00
	Softwoods	0.00	0.86	0.00
	Other hardwoods	0.05	0.07	1.00
B	Boreal hardwoods	0.97	0.00	0.23
	Softwoods	0.00	1.00	0.06
	Other hardwoods	0.03	0.00	0.71

Table 4 Classification matrices using seven different classes (chestnut, oak, insignis pine, scots pine, beech, eucalyptus, and jatobá) obtained by two different nonparametric classification methods, using TG data

Methods	Estimated	Actual						
		Chesn.	Oak	Insig. P.	Scots P.	Beech	Eucal.	Jat.
K-NPFDA	Chestnut	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Oak	0.00	0.72	0.00	0.14	0.00	0.00	0.00
	Insignis P.	0.00	0.00	1.00	0.14	0.00	0.00	0.00
	Scots P.	0.00	0.00	0.00	0.58	0.00	0.00	0.00
	Beech	0.00	0.14	0.00	0.00	1.00	0.00	0.00
	Eucalyptus	0.00	0.14	0.00	0.00	0.00	1.00	0.14
	Jatobá	0.00	0.00	0.00	0.14	0.00	0.00	0.86
B	Chestnut	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Oak	0.00	0.67	0.00	0.00	0.00	0.00	0.00
	Insignis P.	0.00	0.00	1.00	0.00	0.00	0.00	0.00
	Scots P.	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	Beech	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	Eucalyptus	0.00	0.00	0.00	0.00	0.00	1.00	0.17
	Jatobá	0.00	0.33	0.00	0.00	0.00	0.00	0.83

Table 5 Classification matrix using seven different classes (chestnut, oak, insignis pine, scots pine, beech, eucalyptus, and jatobá) obtained by K-NPFDA using a training sample with 42 TG curves

207–307 °C, K-NPFDA	Estimated	Actual						
		Chesn.	Oak	Insig. P.	Scots P.	Beech	Eucal.	Jat.
New sample $P = 1$	Chestnut	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Oak	0.00	1.00	0.00	0.00	0.00	0.00	0.00
	Insignis P.	0.00	0.00	1.00	0.00	0.00	0.00	0.00
	Scots P.	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	Beech	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	Eucalyptus	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	Jatobá	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Probabilities of correct classification of a new sample consisting of seven curves (one per class)

Table 6 Classification matrix using three different classes (boreal hardwoods, softwoods and tropical, and austral hardwoods) obtained by K-NPFDA using a training sample with 42 TG curves

192.5–292.5 °C, K-NPFDA	Estimated	Actual		
		Boreal hardwoods	Softwoods	Other hardwoods
New sample $P = 1$	Boreal hardwoods	1.00	0.00	0.00
	Softwoods	0.00	1.00	0.00
	Other hardwoods	0.00	0.00	1.00

Probabilities of correct classification of a new sample consisting of three curves

Table 7 Correct classification probabilities and optimal intervals obtained by each classification method

Methods	3 Groups classification		7 Groups classification	
	Optimal interval/°C	Corr. class. prob.	Optimal interval/°C	Corr. class. prob.
K-NPFDA	322.5–485.0	0.80	322.5–472.5	0.57
KNN-NPFDA	322.5–485.0	0.80	322.5–347.5	0.60
B	330.0–575.5	0.60	325.0–375.0	0.46
B-PCA	330.0–575.5	0.67	325.0–375.0	0.38

The DSC data were tested using three (boreal hardwoods, softwoods, other hardwoods) and seven classes

Table 8 Classification matrix using three different classes (boreal hardwoods, softwoods, and other hard woods) obtained by different non-parametric classification methods, using DSC curves

Methods	Estimated	Theoretical		
		Boreal hardwoods	Softwoods	Other hardwoods
K-NPFDA	Boreal hardwoods	0.90	0.29	0.21
	Softwoods	0.05	0.71	0.00
	Other hardwoods	0.05	0.07	0.79

In Table 10 we measure the prediction power of a new sample taking a training sample of 42 items. The new sample consists of seven curves, six of which have been

classified successfully ($P = 0.86$). A higher misclassification is obtained precisely in the most heterogeneous group (other hardwoods).

Table 9 Classification matrix using 7 different classes (chestnut, oak, insignis pine, scots pine, beech, eucalyptus, and jatobá) obtained by K-NPFDA method, using DSC data

Methods	Estimated	Actual						
		Chesn.	Oak	Insig. P.	Scots P.	Beech	Eucal.	Jat.
K-NPFDA	Chestnut	0.86	0.14	0.00	0.13	0.00	0.00	0.14
	Oak	0.00	0.57	0.14	0.00	0.14	0.14	0.14
	Insignis P.	0.00	0.00	0.29	0.29	0.14	0.00	0.00
	Scots P.	0.00	0.00	0.43	0.29	0.00	0.00	0.00
	Beech	0.00	0.29	0.14	0.29	0.43	0.00	0.00
	Eucalyptus	0.00	0.00	0.00	0.00	0.00	0.86	0.00
	Jatobá	0.14	0.00	0.00	0.00	0.29	0.00	0.72

Table 10 Classification matrix using three different classes (boreal hardwoods, softwoods and tropical, and austral hardwoods) obtained by K-NPFDA using a training sample with 42 DSC curves

322.5–485.5 °C K-NPFDA	Estimated	Actual		
		Boreal hardwoods	Softwoods	Other hardwoods
New sample $P = 0.86$	Boreal hardwoods	1.00	0.00	0.50
	Softwoods	0.00	1.00	0.00
	Other hardwoods	0.00	0.00	0.50

Probabilities of correct classification of a new sample consisting of three curves

Conclusions

Classifying different species of wood using the TG curves as discriminant characteristic has been proved possible (percentage of correct classification = 90%). Also, the classification between hardwoods, softwoods, and tropical or austral hardwoods have been successfully carried out using these curves (percentage of correct classification = 94%). The results are comparable to those obtained from image-based processing systems and spectrum-based processing systems. It was observed that the temperature ranges corresponding to the higher probabilities of correct classification basically match with those reported for the single components decomposition (mainly hemicellulose).

The DSC curves obtained by a simultaneous SDT have less discriminant power than that of the TG curves. Nevertheless, using these curves, very good results are obtained when we try to distinguish among boreal hardwoods, softwoods, and tropical or austral hardwoods (percentage of correct classification = 80%) and among certain types of hardwoods (chestnut, jatobá, and eucalyptus). Moreover, the referenced temperature range corresponding to the maximum rate of decomposition of lignin and cellulose is the range where more differences among species were found, using DSC curves.

In general, K-NPFDA and KNN-NPFDA methods, based on the nonparametric Nadaraya–Watson functional regression estimator, have provided probabilities of correct

classification superior to the others based on the boosting algorithm, and with a shorter computing time.

Acknowledgements This study was supported by the Ministry of Education and Science MTM2008-00166 (ERDF included), and by Xunta de Galicia, under Grant No.PGIDIT07PXIB105259PR. The authors wish to express special thanks to Manuel Febrero Bande for his valuable comments.

References

1. Guindeo Casasús A, García Esteban L, Peraza Sánchez F, Arriaga Martitegui F. *Especies de maderas*. Madrid: Asociación de investigación técnica de las industrias de la madera yorcho (AITIM); 1997.
2. Lewis IR, Daniel NW, Chaffin NC, Griffiths PR. Raman spectrometry and neural networks for the classification of wood types-I. *Spectrochim Acta A-Mole Biomole Spectrosc*. 1994;50: 1943–58.
3. Miller RB. Structure of wood. In: *Wood handbook: Wood as an engineering material*. Madison, WI: Woodhead Publishing Limited, Department of Agriculture, Forest Service, Forest Products Laboratory; 1999.
4. Cavalin P, Oliveira LS, Koerich AL, Britto AS. Wood defect detection using grayscale images and an optimized feature set. In: *Proceedings IEEE Ind. Electron (IECON)*. Singapore: World Scientific; 2006. pp. 3408–12.
5. Fuentealba C, Simon C, Choffel D, Chawentier P, Massons D. Wood products identification by internal characteristics readings. In: *Proceedings IEEE-ICIT*. Hammamet: IEEE; 2004.
6. Gu IY, Andersson H, Vicen R. Automatic classification of wood defects using support vector machines. In: *Bole L, Kulikowski*

- JL, Wojciechowski K, editors. ICCVG 2008, Lecture Notes in Computer Science. Berlin Heidelberg: Springer-Verlag; 2009. pp. 356–67.
7. Lampinen J, Smolander S, Korhonen M. Wood surface inspection system based on generic visual features. In: Fogelman F, Gallinari SP, editors. Industrial applications of neural networks. Paris: World Scientific; 1995. pp. 35–42.
 8. Watanabe K, Hart JF, Mansfield SD, Avramidis S. Near-infrared technology applications for quality control in wood processing. In: Ridley-Ellis DJ, Moore JR, editors. Proceedings of the final conference of COST Action E53, quality control for wood & wood products. Edinburgh, UK: Forest Products Research Institute/Centre for Timber Engineering, Edinburgh Napier University; 2010.
 9. Khalid M, Lee ELY, Yusof R, Nadaraj M. Design of an intelligent wood species recognition system. *Int J Simul Syst Sci Technol.* 2008;9:9–19.
 10. Lavine BK, Davidson CE, Moores AJ, Griffiths PR. Raman spectroscopy and genetic algorithms for the classification of wood types. *Appl Spectrosc.* 2001;55:960–66.
 11. Nuopponen MH, Birch GM, Sykes RJ, Lee SJ, Stewart DJ. Estimation of wood density and chemical composition by means of diffuse reflectance mid-infrared fourier transform (DRIFT-MIR) spectroscopy. *J Agric Food Chem.* 2006;54:34–40.
 12. Piuri V, Scotti F. Design of an automatic wood types classification system by using fluorescence spectra. *IEEE Trans Syst Man Cybern C-Appl Rev.* 2010;40:358–66.
 13. Yang H, Lewis IR, Griffiths PR. Raman spectrometry and neural networks for the classification of wood types. 2. Kohonen self-organizing maps. *Spectrochim Acta A-Mole Biomol Spectrosc.* 1999;55:2783–91.
 14. Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Berlin: Springer-Verlag.
 15. Ramsay JO, Silverman BW. Functional data analysis 2nd ed. New York, Springer-Verlag; 2005.
 16. Ramsay JO, Silverman BW. Applied functional data analysis. New York: Springer-Verlag; 2002.
 17. Alén R, Kuoppala E, Pia O. Formation of the main degradation compound groups from wood and its components during pyrolysis. *J Anal Appl Pyrolysis.* 1996;36:137–48.
 18. Gašparovič L, Koreňová Z, Jelemenský L. Kinetic study of wood chips decomposition by TGA. *Chem Pap.* 2009;64:174–81.
 19. Grønli MG, Várhegyi G, Blasi C. Thermogravimetric analysis and devolatilization kinetics of wood. *Ind Eng Chem Res.* 2002; 41:4201–08.
 20. Müller-Hagedorn M, Bockhorn H, Krebs L, Müller U. A comparative kinetic study on the pyrolysis of three different wood species. *J Anal Appl Pyrolysis.* 2003;68–69:231–49.
 21. Raveendran K, Ganesh A, Khilar KC. Pyrolysis characteristics of biomass and biomass components. *Fuel.* 1996;75:987–98.
 22. Roberts AF. A review of kinetics data for the pyrolysis of wood and related substances. *Combust Flame.* 1970;14:261–72.
 23. Wang S, Wang K, Liu Q, Gu Y, Luo Z, Cen K, Fransson T. Comparison of the pyrolysis behavior of lignins from different tree species. *Biotechnol Adv.* 2009;27:562–7.
 24. Korošec RC, Lavrič B, Rep G, Pohleven F, Bukovec P. Thermogravimetry as a possible tool for determining modification degree of thermally treated Norway spruce wood. *J Therm Anal Calorim.* 2009;98:189–95.
 25. Mohan D, Pittman JCU, Steele PH. Pyrolysis of wood/biomass for bio-oil: a critical review. *Energy Fuel* 2006;20:848–89.
 26. Bühlmann P, Hothorn T. Boosting algorithms: regularization, prediction and model fitting. *Stat Sci.* 2007;22:477–505.
 27. Prime RB, Bair HE, Gallagher PK, Riga A. Thermogravimetric analysis (TGA). In: Menczel JD, Prime RB, editors. Thermal analysis of polymers Fundamentals and applications. San José, CA: Wiley; 2009. pp. 7–240.
 28. López-Granados F, Peña Barragán JM, Jurado-Expósito M, Francisco-Fernández M, Cao R, Alonso-Betanzos A, Fontenla-Romero O. Multispectral classification of grass weeds and wheat (*Triticum durum*) using linear and nonparametric functional discriminant analysis and neural networks. *Weed Res.* 2008;48: 28–37.
 29. Naya S, Cao R, Artiaga R, García A. New method for material classification from TGA data by nonparametric regression. *Mater Sci Forum.* 2006;514–516:1452–6.
 30. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2008. <http://www.R-project.org>. Accessed 1 Sep 2010.
 31. Fraiman R, Muniz G. Trimmed means for functional data. *Test.* 2001;10:419–40.