

# Classification of wood using differential thermogravimetric analysis

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

Received: 31 July 2014 / Accepted: 18 October 2014 / Published online: 28 November 2014  
© Akadémiai Kiadó, Budapest, Hungary 2014

**Abstract** The aim of this study is to propose an alternative methodology to classify wood species using the first (DTG), second (2DTG), and third (3DTG) derivatives of the thermogravimetric curves (TG). Accordingly, the main contribution of this new procedure consists on classifying materials (wood) taking into account the mass loss rate and acceleration with respect to temperature. In our research, each TG curve is firstly smoothed using the local polynomial regression estimator, and the first, second, and third derivatives are estimated. The application of the local polynomial regression estimator provides a reliable way to obtain the TG derivatives, overcoming the noise problem in the TG derivative estimation. Then, using these estimated curves, the different wood classes are discriminated employing a nonparametric functional data analysis (NPFDA) technique, based on the Bayes rule and the Nadaraya-Watson regression estimator, and also novel functional generalized additive models (GAM).

The latter allows to classify materials using simultaneously more than one type of thermal curves. The results are compared with those obtained using classical and machine learning multivariate supervised classification methods, such as Linear discriminant analysis, Quadratic classification, Naïve Bayes, Logistic regression,  $k$  Nearest neighbors, Neural networks, and Support vector machines. A regression model consisting of the mixture of the first derivatives of four generalized logistic components, one per principal wood constituent (water, hemicellulose, cellulose, and lignin), is applied to fit the DTG curves. The resulting 16 parameters from this fit characterize each curve and are used as datasets to apply the multivariate supervised classification methods. The use of the TG derivatives jointly with the TG curves has proved to be an optimal discriminating feature, when the new functional GAM techniques are employed.

**Keywords** Wood · Supervised classification · Nonparametric functional data analysis · Multivariate analysis · Thermogravimetric analysis · Pyrolysis

---

M. Francisco-Fernández (✉)  
Departamento de Matemáticas, Facultad de Informática,  
Universidade da Coruña, A Coruña, Spain  
e-mail: mariofr@udc.es

J. Tarrío-Saavedra · S. Naya  
Departamento de Matemáticas, Escuela Politécnica Superior,  
Universidade da Coruña, Ferrol, Spain  
e-mail: jtarrío@udc.es

S. Naya  
e-mail: salva@udc.es

J. López-Beceiro · R. Artiaga  
Departamento de Ingeniería Industrial II, Escuela Politécnica  
Superior, Universidade da Coruña, Ferrol, Spain  
e-mail: jorge.lopez.beceiro@gmail.com

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

## Abbreviations

DTG	First derivative of the thermogravimetric curve
2DTG	Second derivative of the thermogravimetric curve
3DTG	Third derivative of the thermogravimetric curve
TG	Thermogravimetric curve
NPFDA	Nonparametric functional data analysis
GAM	Generalized additive models
FTR	Fourier transform Raman
NN	Neural networks
$k$ -NN	$k$ nearest neighbor

SVM	Support vectors machines
PDSC	Pressure differential scanning calorimetry
LDA	Linear discriminant analysis
NBC	Naïve Bayes classifier
FDA	Functional data analysis
K-NPFD	Nonparametric functional data analysis based on kernel methods
IRLS	Iteratively reweighted least-squares
ASE	Average squared error
MDS	Multidimensional scaling
GLM	Generalized linear model

## Introduction

An important task in wood technology is the correct identification of different wood species. This is considered a difficult activity mainly because of the complex structure of this material. In addition, this task is usually very costly because it requires highly qualified personnel. Thus, the application of automatic statistical models for recognition of wood samples is useful and justified. There are some papers addressing the problem of wood identification using different databases obtained with image-based and spectrum-based processing systems, and applying pattern recognition techniques. Non-supervised classification was developed using Fourier transform Raman (FTR) spectroscopy and Neural networks (NN) [1]. Yang et al. [2] applied NN to FTR spectra to distinguish temperate woods from tropical woods. Other supervised classification methods, such as  $k$  nearest neighbor ( $k$ -NN), linear and quadratic discrimination, and support vectors machines (SVM), were employed to classify fluorescence spectra corresponding to wood of different species [3]. The studies of Mallik et al. [4], on the one hand, and Bremananth et al. [5] and Wang et al. [6], on the other hand, represent interesting alternatives based on image segmentation and image textural analysis, respectively. Recently, Rojas et al. [7] used the information provided by stress-wave sounds.

In the context of historical and archeological studies, thermal analysis techniques and kinetic model applications have been recently used to identify wood species from a non-automatic point of view [8, 9]. Using automatic statistical methods, other alternative approaches to image and spectrum processing systems, consisting on applying multivariate and functional statistical techniques to thermogravimetric (TG) and pressure differential scanning calorimetric (PDSC) curves, have been proposed to solve the identification problem [10–12]. However, the possibility of using the derivatives of the TG curves (previously estimated using the local polynomial regression estimator)

as a source of data for statistical classification of wood species has not been studied yet. Thermogravimetry involves analysis of mass or mass percentage versus temperature, and also analysis of mass loss rate versus temperature (DTG). The observation of the DTG curves usually enhances the resolution [13]. This means that DTG curves are useful to distinguish overlapped mass loss events, to identify trend and maximum rate of mass loss processes, and also to identify small mass loss steps [14]. These features could make the DTG curves a successful tool to classify materials such as wood, with degradation processes of its main components (cellulose, hemicellulose, and lignin) strongly overlapped [10, 11, 15–17]. However, the process of differentiation applied to obtain the DTG curves normally produces an increase of noise. Therefore, data are usually preprocessed (by smoothing or fitting a parametric or a nonparametric model) to estimate the corresponding derivative [13]. In the present research, we tackle this problem, focusing not only on the first TG derivative (DTG), but also on the second (2DTG) and third (3DTG) ones. Specifically, the present work should give an answer to the following questions: Is it possible to observe differences among wood species in the mass loss rate of wood samples? and Is it possible to classify woods from the differences present in the TG curve derivatives?

Wood degradation in an inert atmosphere is dominated by the degradation behavior of its three main components [18]: cellulose, lignin, and hemicellulose [16, 18–24]. Cellulose represents about 40–60 % in the mass of dry wood (being 23–33 % in the case of softwoods), about 23–33 % is lignin in softwoods, being 16–25 % in the case of hardwoods, and finally 25–35 % is hemicellulose (more in hardwoods than in softwoods) [1, 22, 25]. The TG or the DTG curves, describing the pyrolysis of wood, nearly coincide with the sum of its constituent degradation [19, 24]. Moreover, the proportion of each wood component varies, to a greater or lesser extent, depending on the species [16, 20, 22, 23]. Therefore, we can guess the potential of these curves to classify different wood species. In this framework, we can stress the work of Dai et al. [26], where a procedure for the kinetic analysis on the decomposition of lignocellulosic materials is proposed.

It is important to note that there are no studies about wood supervised classification using the derivatives of the TG curves (in many cases, more informative and useful than the TG curves). In fact, thermal analysis techniques are not usual in automatic classification tasks. One of the most important reasons is the relative difficulty to deal with the information given by a curve of infinite dimension. The most common way to overcome this difficulty (present in many problems, corresponding to diverse science fields) is the reduction of the curve dimension. However, following this approach, a new problem arises; What features are

actually representative of each curve? In the present study and regarding the DTG curves, a nonlinear regression model, related to logistic mixture regression [10], is proposed to explain the mass loss rate with respect to the increasing temperature. It consists of four components, derivatives of the generalized logistic function, one per principal constituent of wood (hemicellulose, cellulose, and lignin), and one corresponding to the water mass lost rate of the sample. The obtained regression parameters (four for each component, 16 altogether) are used as a vector of features, summarizing properly the information given by each curve. Using this procedure, it is possible to apply multivariate supervised classification methods, such as linear discriminant analysis (LDA), Logistic Regression, Naïve Bayes (NBC),  $k$ -NN, SVM, or NN.

Some statistical techniques designed to be applied to functional data (named functional data analysis (FDA) techniques) have been studied in the last years. They allow to work with infinite dimensional data (curves or functions) without having to reduce their dimension. The interest in FDA methods has increased considerably, because the technological progress allows collecting observations of infinite dimensional objects. Some references describing some of the most popular functional data analysis approaches are [27–36]. In the present study, we compare nonparametric functional data analysis (NPFDA), functional generalized additive models (GAM), and classical multivariate statistical approaches to classify wood samples from their TG derivatives (TG curves and their derivatives up to order three). Considering the functional nature of the curves used here and the generally good performance of the statistical nonparametric and semi-parametric methods in other fields, it is expected that NPFDA and GAM procedures will provide good results in this framework.

Accordingly, the objectives of this study are

1. Evaluating the potential of the first DTG curves, second and third TG derivatives as a source of information for wood classification tasks.
2. Introducing and applying a reliable smoothing method to overcome the problem of noise in the TG derivative estimation: the local polynomial regression estimator.
3. Implementing and evaluating the behavior of functional classification techniques. Special attention is paid to GAM models, because they represent a novel alternative that enables the use of more than one type of curves (e.g., TG and DTG curves simultaneously) to perform wood classification.
4. Comparing the results obtained by NPFDA and GAM supervised classification techniques with those obtained using multivariate methods, when the TG and the DTG curves are used. For this last purpose, a

parametric regression model to summarize a curve information (infinite dimension) using 16 parameters (finite dimension) was proposed.

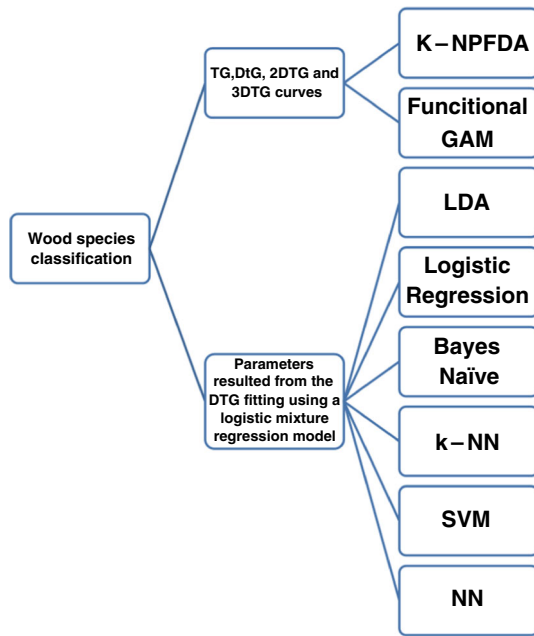
5. Finding the temperature intervals where the highest probability of correct classification is reached, when using the NPFDA technique with the DTG curves and their first and second derivatives. Relating these results with the degradation temperature ranges corresponding to the main wood constituents.

## Experimental

Seven samples corresponding to seven different wood species of industrial interest are tested: five hardwoods (beech or *Fagus sylvatica*, chestnut or *Castanea sativa*, european oak or *Quercus robur*, jatobá or *Hymenaea courbaril*, and *Eucalyptus globulus*) and two softwoods (Scots pine or *Pinus sylvestris* and insignis pine or *Pinus radiata*). Seven samples per each species are selected. They are obtained from wood in the inner part of the trunk of different commercial planks. The sample mass tested is between 6 and 8 mg. Although the sample size used in this experiment is not very large, from our previous experience, we can deduce that if the samples are obtained from inside the trunk and are well dried (as it is the case here), the results are very similar independently from other possible factors. Therefore, we can assume that the samples considered in the present research represent quite well the species. The aim of this sampling process is to obtain a compromise between capturing the existing variability and minimizing the time of experimentation. The thermogravimetric tests were performed in a TA SDT 2960 TA Instruments simultaneous analyzer. The experimental setup consisted of a 20 °C min<sup>-1</sup> linear heating from room temperature to 600 °C. Open alumina crucibles were used, and a 50 mL min<sup>-1</sup> flow of nitrogen was kept along the experiments. The nitrogen was purged for 10 min, before starting the heating program for establishing an inert environment. That high heating rate is implemented to obtain a no time-consuming classification procedure, and also to test the classification methodologies in the worst conditions.

## Statistical analysis: classification and regression methods

Supervised classification can be defined as a procedure assigning an instance (a finite dimension vector or a infinite dimension curve) to a certain group out of several possible alternatives. This is a process with multiple applications in



**Fig. 1** Classification methods and datasets

industry, chemometrics, and materials science in general. The proposed classification methods try to find a relationship between the explanatory variables (multidimensional vectors or functional data, if it is the case) and a class indicator or group. The aim is to predict correctly the class of a new sample. This paper addresses this problem from two perspectives: functional and multivariate analysis (see Fig. 1). The free statistical software R [37] was employed to implement the multivariate and the nonparametric functional methods used in the present paper.

#### NPFDA classification methods

A nonparametric functional technique based on kernel methods (K-NPFDA) is applied to construct a classification rule to discriminate between the different wood species [32]. A sample of 49 thermal analysis curves, including the TG curves and, in general, their derivatives up to order 3 are used for this task. Using this approach, a curve is classified as belonging to the species or the group to which the highest posterior probability is obtained. The functional Nadaraya-Watson kernel nonparametric method, shown in (1), is applied. Given a new curve,  $x = x(t)$ , obtained from a material (wood, in this case) to classify, the estimator of the posterior probability of belonging to a class  $g$ , with  $g \in \{0, 1, \dots, G\}$ , is given by

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

where the observed 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. Note that in the present study, the sample  $X_i$ ,  $i = 1, \dots, n$ , can represent the TG curves or their first (DTG), second, or third derivatives.

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. 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:

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

As pointed out previously, several scripts using the free statistical software R were written implementing these nonparametric approaches. Additionally, some functions belonging to `fda.usc` and `fda` R libraries [37, 38] were also used.

#### Functional GAM classification method

This classification is one of the most recent classification methods developed in the FDA field. Thus, this is the first time that it is applied to materials science and thermal analysis.

The categorical response variable  $y$  is estimated through a sum of  $k$  smooth functions of the covariates  $X$  and a link function [39].

$$E(y) = \mu = g^{-1} \left( \beta_0 + \sum_{j=1}^k f_j(X_j) \right),$$

with  $X_j$  the columns of  $X$  and  $E(f_j(X_j)) = 0$ . The estimation of the model parameters is performed using the IRLS algorithm with backfitting steps [40]. The  $\beta$  parameters and functions of the covariates,  $f_j(X_j)$ , are expressed using  $b$ -splines basis. For more information, consult the work of Febrero-Bande and González-Manteiga [39].

In the present case, the  $y$ -variable represents the wood species, and the  $X_j$ -covariates can be the TG, DTG, 2DTG, and 3DTG curves. More than one covariate or type of curves can be used simultaneously in the model to estimate the wood class.

As in the case of the K-NPFDA method, we have written some code in R to implement this procedure, also using functions of the `fda.usc` R library.

Multivariate classification techniques

An alternative approach to the functional method described in the previous section consisted in discretizing the curves, selecting some features of them, and applying a multivariate classification technique. Following these lines, classical multivariate methods and machine learning approaches were used to classify the wood samples. To perform this analysis, we focused on the DTG curves. A new technique for DTG curve discretization was considered. This discretization was done using the parameters resulting from the fit of a nonlinear parametric model to the DTG curves. The model proposed to fit the experimental data consisted of a four term mixture of the first derivative of a generalized logistic function. These terms are related to the main wood components: cellulose, hemicellulose, lignin, and water:

$$x(t) = \sum_{i=1}^n \frac{c_i \times b_i \times \exp(-b_i \times (t - m_i))}{[1 + \tau_i \times \exp(-b_i \times (t - m_i))]^{(1+\tau_i)/\tau_i}}, \tag{2}$$

where the  $c$  parameter represents the mass involved in the degradation process,  $b$  is related to the decomposition rate or rate of change,  $\tau$  accounts for the asymmetry,  $m$  represents the temperature at the maximum rate of change, and  $t$  is the temperature. The optimal fits were obtained by minimizing the average squared error (ASE). Once these fits and the discretization process were carried out, several multivariate classification methods, such as LDA, Logistic Regression, NBC,  $k$ -NN, SVM and NN, were applied. Next, a brief description of these approaches is presented.

The first multivariate classification technique, LDA, was proposed by Fisher [41]. This method assumes multivariate normality and equal covariance matrices between groups. In the context of wood classification, LDA was applied in [4].

The logistic model can be employed when the basic LDA assumptions are not fulfilled [42]. In the context of the present paper, it is important not to confuse the Logistic Regression classifier with the regression model composed of four logistic terms used to characterize the DTG curves.

The logit model is generalized to more than two populations, i.e., for qualitative response with more than two possible levels. Denoting by  $G$  the number of different populations, and  $p_{ig}$  the posterior probability that an observation  $i$  belongs to the class  $g$ , then

$$p_{ig} = \frac{\exp(\beta_{0g} + \beta'_{1g}x_i)}{1 + \sum_{j=1}^{G-1} \exp(-\beta_{0j} - \beta'_{1j}x_i)}$$

The NBC classifier is based on the Bayes rule. This method selects the event class or the sample as that corresponding to the largest posterior obtained probability (that is, the probability that a sample,  $X = \{x_1, x_2, \dots, x_d\}$ , belongs

to a particular class,  $C_j$ , from a group of possible classes  $C = \{C_1, C_2, \dots, C_k\}$ ). Note that assuming independence of the variables, the posterior probabilities are the following:

$$p(C_j|X) \propto p(C_j) \prod p(x_k|C_j)$$

The  $k$ -NN procedure is a well-known multivariate non-parametric supervised classification method [43]. Any parametric assumption has to be supposed for its application. A description of this approach can be found in [10].

SVM is a machine learning multivariate method, where the classification is done by constructing hyperplanes in a multidimensional space that separates the cases with different class labels [44]. See, for example, [10] for a description of this method.

NN is a nonlinear machine learning model based on the structure and performance of biological neural networks [45]. It consists of an interconnected group of artificial neurons grouped in layers that process information. In the special case of supervised classification, the mapping implied by the data is inferred and the cost function, related to the mismatch between the developed mapping and the data, is calculated.

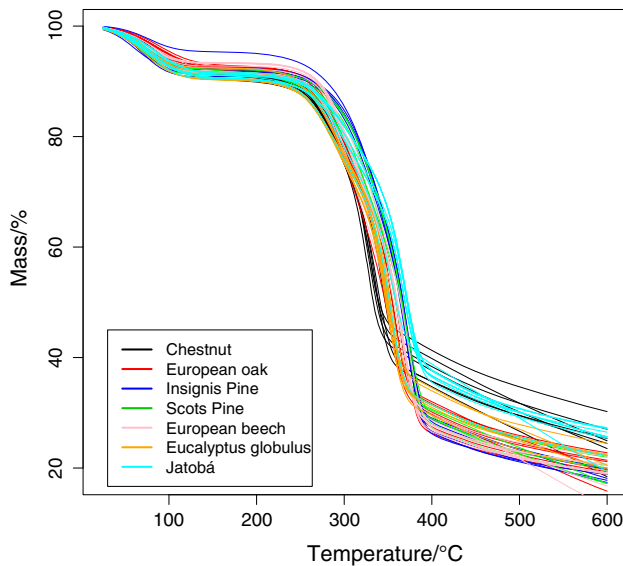
In our study, some functions belonging to `e1071`, `nnet`, `MASS`, and `kknn` R libraries [37] were used to implement the previous multivariate classification methods.

Results and discussion

This section presents the application of the methods described in Sects. [NPFDA classification methods](#), [Functional GAM classification method](#), and [Multivariate classification techniques](#). Different thermal analysis curves of a sample of 49 wood items to classify them in different species were used. The procedure applied to validate the proposed classification methods was a double leave-one-out cross-validation technique. It works by leaving out one curve; then a model is trained with the 48 remaining curves (using another leave-one-out process), and finally, the model obtained is used for classifying the curve left out. This is repeated until all the curves have been left out once [32, 46]. Finally, the proportion of success or probability of correct classification corresponding to each classification method is calculated (overall and for each species). The leave-one-out cross-validation procedure is specially suitable when the number of samples is small. This validation procedure can be consulted in [10] or in [47].

Figure 2 shows the 49 TG traces corresponding to the seven different species, seven per species, after smoothing these curves with the nonparametric local linear estimator [48]. Note that each legend in this figure refers to the seven TG curves of each species. The local linear estimator is a





**Fig. 2** The 49 smoothed TG curves using the linear local polynomial estimator, corresponding to the 49 samples tested, seven per species (each *legend* refers to the seven TG curves of each species)

particular case of the more general local polynomial estimator. This is a well-known nonparametric procedure used to estimate the mean function (or the trend) in a regression problem. In the framework of the present paper, denoting by  $\{(t_k, Y_k)\}_{k=1}^n$ , the observed values of a specific TG curve,  $Y_k$ , for several temperature levels,  $t_k$ , with  $k = 1, 2, \dots, n$ , the following regression model could be formulated:

$$Y_k = m(t_k) + \varepsilon_k, \quad 1 \leq k \leq n, \tag{3}$$

where  $\varepsilon_k$  are random errors. Our aim is “smoothing the data” or, in other words, estimating the mean function  $m(\cdot)$ . With this smoothing process, we seek to eliminate the noise of the experiment, i.e., all the variability present in the curve that does not correspond to the experimental conditions scheduled. Thus, the TG curves become an almost functional relationship between mass and temperature.

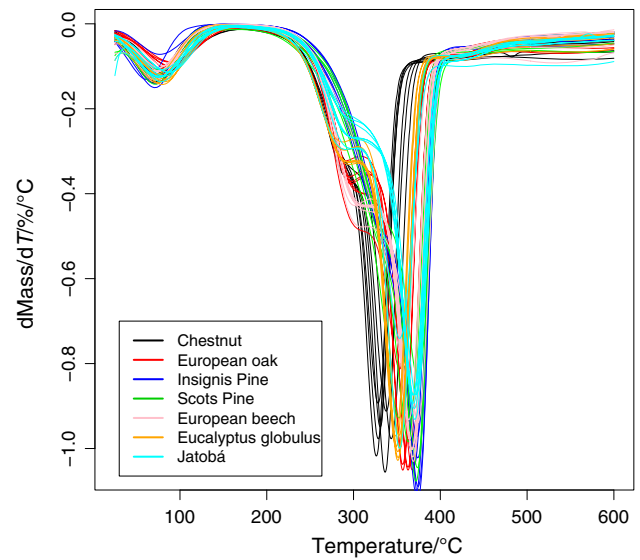
The local polynomial regression estimator consists in locally fitting a  $p$ -degree polynomial to the observed sample,  $\{(t_k, Y_k)\}_{k=1}^n$ , and it can be written as

$$\hat{m}_h(t) = e_1' (X_t' W_t X_t)^{-1} X_t' W_t Y \tag{4}$$

where  $e_1 = (1, 0, \dots, 0)'$ ,  $Y = (Y_1, \dots, Y_n)'$ ,

$$X_t = \begin{bmatrix} 1 & t_1 - t & \dots & (t_1 - t)^p \\ \vdots & \vdots & & \vdots \\ 1 & t_n - t & \dots & (t_n - t)^p \end{bmatrix},$$

and  $W_t = \text{diag}\{K((t_1 - t)/h), \dots, K((t_n - t)/h)\}$ , with  $K(\cdot)$  a kernel function,  $p$  the degree of the local polynomial (the local linear estimator corresponds to the case of

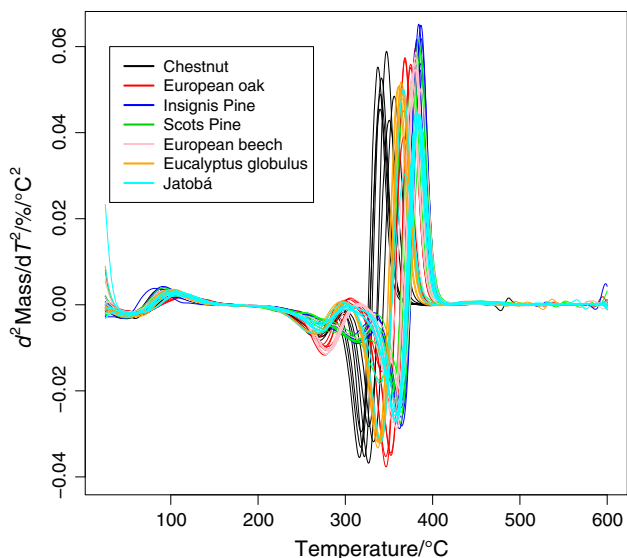


**Fig. 3** Estimations of the 49 DTG curves or TG derivatives, using the linear local polynomial estimator, corresponding to the 49 samples tested, seven per species (each *legend* refers to the seven DTG curves of each species)

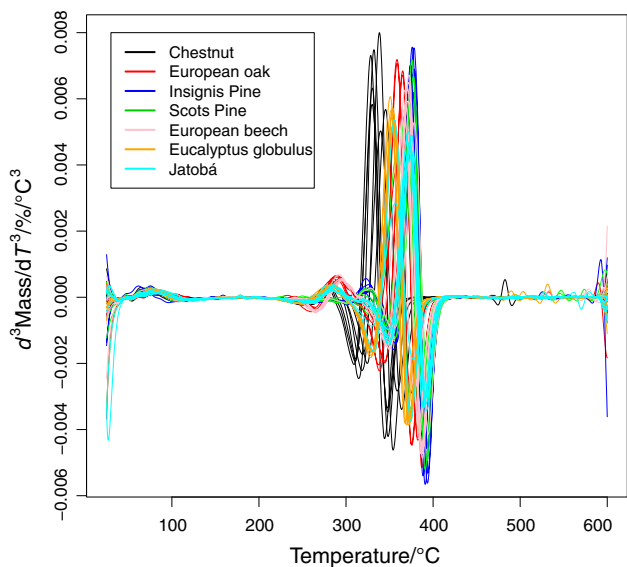
$p = 1$ ), and  $h$  the bandwidth. Similarly to the estimator given in (1), the bandwidth  $h$  is a very important parameter to be selected by the user in order to obtain reliable estimators. In Fig. 2, the corresponding bandwidth  $h$  was selected using a direct plug-in method [49].

An important advantage of the local polynomial regression estimator is that estimators of the derivatives of the mean function can be directly obtained [48]. For example, Fig. 3 presents the estimators of the first derivative of the TG curves (or DTG curves), calculated using the nonparametric local polynomial model, with a second degree polynomial, applied to the TG data. It can be observed in Fig. 3 that the DTG signal noise, usually much higher when numerical derivation is used than that in the TG signal, has been attenuated to a large extent. Additionally, estimators of the DTG first and second derivatives, obtained from the nonparametric model, are shown in Figs. 4 and 5, respectively. Local polynomials of degrees 3 and 4, respectively, were used to perform these fits. Similarly in Fig. 2, each legend in Figs. 3, 4, and 5 refers to the corresponding derivatives of the seven TG curves of each species.

In a first step, the estimated DTG curves were used to classify the wood species, applying FDA classification methods (K-NPFDA and GAM models), described in Sects. [NPFDA classification methods](#) and [Functional GAM classification method](#), respectively, and the multivariate approaches, presented in Sect. [Multivariate classification techniques](#). Regarding the K-NPFDA technique, we have observed that a prior smoothing of the data (with the local

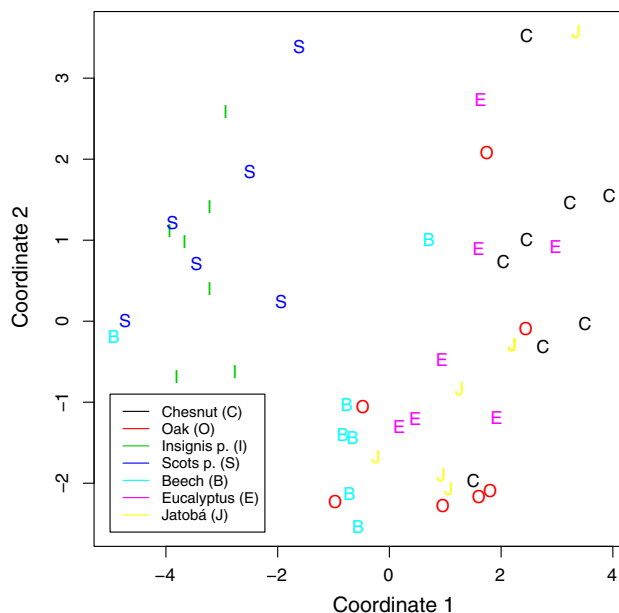


**Fig. 4** Estimations of the second derivatives of the 49 TG curves, using the linear local polynomial estimator, corresponding to the 49 samples tested, seven per species (each *legend* refers to the seven 2DTG curves of each species)



**Fig. 5** Estimations of the third derivatives of the 49 TG curves, using the linear local polynomial estimator, corresponding to the 49 samples tested, seven per species (each *legend* refers to the seven 3DTG curves of each species)

linear estimator and an optimal bandwidth) to estimate the DTG curves led to a better correct classification percentage than that obtained when the DTG curves directly provided by the analyzer were used. Specifically, a probability of correct classification of 78 %, using the rough data, and a 82 %, using the estimated DTG curves. On the other hand, regarding the multivariate classification, after adjusting the



**Fig. 6** Samples from different wood species, using MDS with two principal coordinates

parametric model consisting of the mixture of the first derivatives of four generalized logistic components, the resulting parameters were scaled, that is, each column of the feature matrix was centered and divided by its standard deviation for matching the feature variabilities. Then, the metric multidimensional scaling (MDS) technique was applied to obtain descriptive information about the multivariate data [42]. The MDS method calculates the distance matrix between samples. The information provided by this matrix can be approximated by two principal coordinates corresponding to the two largest eigenvalues of the matrix [42]. In Fig. 6, the data structure is plotted using two coordinates. It shows that the samples belonging to the same species tend to be closer than the samples belonging to different ones. All the species samples are grouped. The results seem good, but there is some confusion between Scots pine and insignis pine and between oak and beech. At a first glance, it seems that the parameters properly summarize the information of each curve, being useful to identify and classify these wood species. However, this descriptive study has to be corroborated with the results provided by the supervised multivariate classification methods.

Tables 1 and 2 show the percentages of correct classification when the different approaches are used. Table 1 is referred to the FDA classification methods (K-NPFDA and GAM models), and Table 2 shows the results of the multivariate classification procedures. Note that when classifying among the seven wood species with the estimated DTG

**Table 1** Prediction probabilities for seven different classes obtained by each FDA classification method, using the smoothed TG curves, and their estimated derivatives up to order 3

Methods	TG Pred.	DTG Pred.	2DTG Pred.	3DTG Pred.
K-NPFDA	0.78	0.82	0.84	0.90
GAM	0.78	0.80	0.76	0.73

**Table 2** Prediction probabilities for seven different classes obtained by each multivariate classification method. The multivariate methods are applied using the extracted features from the TG and the DTG curves

Classification methods	TG Predict.	DTG Predict.
LDA	0.76	0.90
Logistic regression	0.76	0.92
NBC	0.65	0.78
<i>k</i> -NN	0.71	0.78
SVM	0.65	0.92
NN	0.73	0.90

curves, using both functional and multivariate methods, the results are excellent. However, the multivariate methods seem to perform better than the functional approaches, especially SVM using a Gaussian kernel (0.92), Logistic Regression (0.92), LDA (0.90), and NN (0.90). In general, results in Table 2 show that it is possible to classify among the seven studied species using these 16 parameters, obtained from the parametric fits as original multivariate data, jointly with the proposed classification methods. We can say that there are differences in the mass loss rate of the wood species and that the 16 parameters corresponding to each DTG curve summarize correctly these differences. It is important to note that LDA and Logistic Regression methods require a shorter computing time.

Tables 1 and 2 also show, in the first column, the results when the smoothed TG curves were used. The best results were obtained with the K-NPFDA and the GAM methods (0.78, in both cases), slightly worse than using the estimated DTG curves with these techniques (0.82 and 0.80, for the K-NPFDA and the GAM procedures, respectively). In this case, to apply the multivariate classification approaches, the expression of the model of generalized logistic derivatives, given in (2), was integrated, obtaining a parametric model for the smoothed TG curves. In view of the results presented in Tables 1 and 2, the probabilities of correct classification using the DTG curves are always significantly higher than those obtained for the TG curves. In fact, as above mentioned, the DTG curves can provide higher resolution than TG curves and thus more information to classify the different wood species.

Functional data classification can also be done from the successive derivatives of the DTG curves (see Table 1). The probability of correct classification obtained from the estimated DTG second derivatives, applying the K-NPFDA method, is particularly high (0.90) and competitive with the results obtained using the multivariate methods with the DTG curves. It is interesting to observe that, in this case, the 3DTG or mass loss acceleration rate resulted more informative (0.90) than the DTG curves and their first derivatives (0.82 and 0.84, respectively). In addition, the 2DTG or mass loss acceleration is more informative than the DTG curve, while the DTG curves give more information about the wood species than the TG curves in this classification problem.

Functional techniques present the advantage of not having to perform any regression model to obtain the corresponding parameters. Note that we have not adjusted parametric models to the first and second derivatives of the DTG curves due to the increasing difficulty in terms of physical interpretation and the required number of components.

The results in Tables 1 and 2 are an indication of the potential interest that may have the study of mass loss acceleration and mass loss acceleration rate curves in the field of thermal analysis; in particular, in the case of material identification. The use of the local polynomial regression estimator to smooth the data allows to estimate the derivatives of the TG curves avoiding numerical derivation. As a result, much of the uncontrolled variability or noise, characteristic of the successive derivatives of TG curves, can be removed.

The analysis of the K-NPFDA method was completed detecting the intervals of the temperature range where this method provided better results. To do this, 1,280 intervals of temperatures were considered and the K-NPFDA procedure was applied in each one of them. The highest probabilities of correct classification were obtained in the temperature ranges of 209.1–305.1, 139.6–331.4, and 163.6–307.5 °C, when using the estimated DTG curves, and their estimated first and the second derivatives, respectively. The hemicellulose, cellulose, and lignin, the three main wood components, decompose at temperatures ranging between 200–260, 240–350, and 280–500 °C, respectively [16, 18, 20, 22, 23, 25]. Therefore, it can be observed that the referenced temperature ranges corresponding to the decomposition of hemicellulose and cellulose are the ranges where more differences among species were found, when using the estimated DTG curves and their first and second derivatives.

Regarding the functional GAM model, as previously mentioned, this is a very recent FDA classification method that can perform identification of materials using more than one type of curves. This is the only method that is able to



**Table 3** Prediction probabilities for seven different classes obtained by each GAM classification method, using the smoothed TG curves, and their estimated derivatives up to order 3

Methods	TG + DTG Pred.	TG + 2DTG Pred.	TG + 3DTG Pred.	DTG + 2DTG Pred.
GAM	<b>0.92</b>	0.90	<b>0.92</b>	0.84

work with more than one type of curves or functional covariates, in addition to the functional generalized linear model (GLM) approach [39]. Table 3 shows the results for the GAM method when it is applied to the curve pairs TG-DTG, TG-2DTG, TG-3DTG, and DTG-2DTG in the interval between 200 and 400 °C. The best results are highlighted in bold in Table 3. This is the range corresponding to the hemicellulose and cellulose thermal degradation.

Comparing the results in Tables 2 and 3, it can be observed that higher proportions of good classification are obtained for the GAM approach when two types of curves are used. Thus, the TG derivatives actually add information that the TG curves did not provide. Due to the limited degrees of freedom (49 curves), the use, as covariates, of three or more type of curves is not possible. It is important to note that the GAM model provides the highest proportion of correct classification, 0.92, using the TG and DTG

**Table 4** Confusion matrix or probabilities of correct classification in seven groups, obtained with the second derivatives of the estimated DTG curves when applying the K-NPFDA method, and using the parameters extracted from the nonlinear fitting model of the estimated DTG curves when applying SVM and Logistic regression. The confusion matrix obtained applying the functional GAM method to the TG and DTG curves simultaneously is also presented

Methods	Estimated	Actual						
		Chestn.	Oak	Insig. P.	Scots P.	Beech	Eucal.	Jat.
K-NPFDA (from DTG 2nd derivative)	Chestn.	0.86	0.00	0.00	0.00	0.00	0.00	0.00
	Oak	0.00	0.57	0.00	0.00	0.00	0.00	0.00
	Insig. P.	0.00	0.00	1.00	0.14	0.00	0.00	0.00
	Scots P.	0.00	0.00	0.00	0.86	0.00	0.00	0.00
	Beech	0.00	0.29	0.00	0.00	1.00	0.00	0.00
	Eucal.	0.00	0.14	0.00	0.00	0.00	1.00	0.00
	Jat.	0.14	0.00	0.00	0.00	0.00	0.00	1.00
SVM (from DTG curves)	Chestn.	1.00	0.00	0.00	0.00	0.00	0.00	0.14
	Oak	0.00	0.86	0.00	0.00	0.00	0.00	0.00
	Insig. P.	0.00	0.00	1.00	0.29	0.00	0.00	0.00
	Scots P.	0.00	0.00	0.00	0.71	0.00	0.00	0.00
	Beech	0.00	0.14	0.00	0.00	1.00	0.00	0.00
	Eucal.	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	Jat.	0.00	0.00	0.00	0.00	0.00	0.00	0.86
Log. Reg. (from DTG curves)	Chestn.	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Oak	0.00	0.86	0.00	0.00	0.00	0.00	0.00
	Insig. P.	0.00	0.00	1.00	0.29	0.00	0.00	0.00
	Scots P.	0.00	0.00	0.00	0.71	0.00	0.00	0.14
	Beech	0.00	0.14	0.00	0.00	1.00	0.00	0.00
	Eucal.	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	Jat.	0.00	0.00	0.00	0.00	0.00	0.00	0.86
GAM (from TG and DTG curves)	Chestn.	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
	Insig. P.	0.00	0.00	0.71	0.14	0.00	0.00	0.00
	Scots P.	0.00	0.00	0.29	0.86	0.00	0.00	0.00
	Beech	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	Eucal.	0.00	0.00	0.00	0.00	0.00	1.00	0.14
	Jat.	0.00	0.00	0.00	0.00	0.00	0.00	0.86

as covariates. The addition of DTG functional covariate increases in almost 15 % the proportion of good classification. When the pair DTG-2DTG is used, the results are slightly lower because of the increasing of noise in derivative estimation. The application of functional GAM classification models could be even more useful when applied to larger databases. This method is strongly recommended for timber classification.

Table 4 shows the confusion matrix corresponding to the seven species, using the best multivariate and functional methods. Looking at the diagonal of this table, all the particular class probabilities are very high. There is only a little of confusion between oak and beech, and between Scots and insignis pine, possibly because the physical properties of these two pairs of wood are very similar. In the case of K-NPFDA method, the confusion between beech and oak is relatively high, but, apart from this case, there is almost no confusion. In the case of functional GAM, SVM, and Logistic Regression methods, the largest confusion occurs between the two types of pine species, although this is relatively small. This confusion obeys to its similar thermal degradation.

## Conclusions

The obtained probabilities of correct classification show that it is possible to observe differences among wood species studying thermal stability tested by DTG. Thus, DTG is an appropriate technique for classifying wood species.

The functional GAM model is a flexible and useful proposal to classify wood species. The highest proportion of correct classification, 0.92, is obtained by this method. This is a high value compared to the results provided by other methodologies in the wood identification field. The GAM model permits to work simultaneously with TG curves and their derivatives. This represents an advance with respect to other approaches.

The functional nonparametric K-NPFDA classification method has also been successfully applied to discriminate among wood species. The best results were obtained when the second derivatives of the DTG curves (0.90), related to the wood mass loss rate, were used.

The application of local polynomial smoothing with an optimal bandwidth significantly improved the correct classification percentages. It removed much of the uncontrolled variability or characteristic noise of the successive TG curve derivatives. As a consequence of this, we observed that the DTG second derivative or mass loss acceleration rate provided more information than the rest of the thermogravimetric curves considered for classification purposes. In addition, the DTG first derivative or mass loss acceleration was more informative than the DTG curves,

while the DTG curves gave more information than the TG traces, in this supervised classification problem.

Mainly, the referenced temperature intervals corresponding to the decomposition of hemicellulose and cellulose were the ranges where more differences among species were found, using the DTG curves and their first and second derivatives, when the K-NPFDA method was used.

A parametric regression model consisting of the mixture of the first derivatives of four generalized logistic components to fit the DTG curves and to extract some features of these curves was proposed, obtaining very good performance. The model is justified on the basis that there are mainly four different degradation processes. Using these features (represented by 16-dimensional vectors), multivariate classifiers such as SVM with Gaussian kernel (0.92), NN (0.90) and Logistic Regression (0.92) produced slightly better probabilities of correct classification than functional methods when the DTG curves were used. However, the K-NPFDA approach was able to give similar results (with a shorter computing time) when the estimated DTG second derivatives were employed.

In summary, functional nonparametric and semiparametric classification techniques (GAM and K-NPFDA) are competitive if they are compared with traditional and machine learning multivariate procedures. FDA methods use all the information provided by the thermal curves and can be run more quickly, avoiding the data dimension reduction process. This makes GAM and K-NPFDA functional methods appropriate techniques to classify wood species using thermogravimetric data.

This paper shows a first approximation to classify wood species using the derivatives of the TG curves and FDA techniques. Although the sample size employed is not very large, we can assume (based on our previous experience) that the samples considered represent quite well the species. Of course, in a future research, a more intensive sampling would be possible to study the influence of several factors on the intra-species variability. Additionally, this work establishes a starting point that opens the possibility of applying these procedures to other materials.

**Acknowledgements** This research has been supported by the Spanish Ministry of Science and Innovation, Grant MTM2008-00166 (ERDF included) and Grant MTM2011-22393. The authors wish to express special thanks to Manuel Oviedo for his valuable comments.

## References

1. Lewis IR, Daniel NW, Chaffin NC, Griffiths PR. Raman spectroscopy and neural networks for the classification of wood types-1. *Spectrochim Acta A, Mol Biomol Spectrosc.* 1994;50:1943–58.
2. Yang H, Lewis IR, Griffiths PR. Raman spectroscopy and neural networks for the classification of wood types. 2. Kohonen self-organizing maps. *Spectrochim Acta Part A.* 1999;55:2783–91.

3. Piuri V, Scotti F. Design of an automatic wood types classification system by using fluorescence spectra. *IEEE Trans Syst Man Cybern Part C-Appl Rev.* 2010;40:358–66.
4. Mallik A, Tarrío-Saavedra J, Francisco-Fernández M, Naya S. Classification of wood micrographs by image segmentation. *Chemometrics Intell Lab Syst.* 2011;107:351–62.
5. Bremananth R, Nithya B, Saipriya R. A wood species recognition system. *Int J Intell Syst Technol.* 2009;4:54–60.
6. Wang HJ, Zhang GQ, Qi HN. Wood recognition using image texture features. *PLoS One.* 2013;8(e76):101.
7. Rojas JAM, Alpuente J, Postigo D, Rojas IM, Vignote S. Wood species identification using stress-wave analysis in the audible range. *Appl Acoust.* 2001;72:934–42.
8. Budrugaec P, Emandi A. The use of thermal analysis methods for conservation state determination of historical and/or cultural objects manufactured from lime tree wood. *J Therm Anal Calorim.* 2010;101:881–6.
9. Cavallaro G, Donato DI, Lazzara G, Milioto S. A comparative thermogravimetric study of waterlogged archaeological and sound woods. *J Therm Anal Calorim.* 2011;104:451–7.
10. Francisco-Fernández M, Tarrío-Saavedra J, Mallik A, Naya S. A comprehensive classification of wood from thermogravimetric curves. *Chemometrics Intell Lab Syst.* 2012;118:159–72.
11. Tarrío-Saavedra J, Naya S, Francisco-Fernández M, López-Beceiro J, Artiaga R. Functional nonparametric classification of wood species from thermal data. *J Therm Anal Calorim.* 2011;104:87–100.
12. Tarrío-Saavedra J, Francisco-Fernández M, Naya S, López-Beceiro J, Artiaga R. Wood identification using pressure DSC data. *J Chem.* 2013;27:475–87.
13. Wendlandt WW, Gallagher PK. Instrumentation. In: Turi EA, editor. *Recent advances in functional data analysis and related topics.* New York: Academic Press; 1981.
14. 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: Wiley; 2009.
15. Diblasi C. Modeling chemical and physical processes of wood and biomass pyrolysis. *Prog Energy Combust Scil.* 2008;34:47–90.
16. Gronli MG, Várhegyi G, Blasi C. Thermogravimetric analysis and devolatilization kinetics of wood. *Ind Eng Chem Res.* 2002;41:4201–8.
17. Sebio-Puñal T, Naya S, Lopez-Beceiro J, Tarrío-Saavedra J, Artiaga R. Thermogravimetric analysis of wood, holocellulose, and lignin from five wood species. *J Therm Anal Calorim.* 2012;109:1163–7.
18. 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.
19. Gašparovič L, Koreňová Z, Jelemenský L. Kinetic study of wood chips decomposition by TGA. In: *Proceedings 36th international conference of SSCHE. Tatranské Matliare, World Scientific; 2009, vol 178, pp. 1–14.*
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. Yang H, Yan R, Chen H, Lee DH, Zheng C. Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel.* 2007;86:1781–8.
25. Mohan D, Pittman JCU, Steele PH. Pyrolysis of wood/biomass for bio-oil: a critical review. *Energy Fuel.* 2006;20:848–89.
26. Dai J, Liu N, Shu L. Applicability of bi-pseudo component separated-stage model for decomposition of lignocellulosic materials in air at multiple heating rates. *J Therm Anal Calorim.* 2011;104:983–90.
27. Cuevas A, Febrero M, Fraiman R. Linear functional regression: the case of fixed design and functional response. *Can J Stat.* 2002;30:285–300.
28. Cuevas A, Febrero M, Fraiman R. An anova test for functional data. *Comput Stat Data Anal.* 2004;47:111–22.
29. Ferraty F, Romain Y. *The Oxford handbook of functional data analysis.* Oxford: Oxford University Press; 2010.
30. Ferraty F, Vieu P. The functional nonparametric model and application to spectrometric data. *Comput Stat.* 2002;17:545–64.
31. Ferraty F, Vieu P. Curves discrimination: a nonparametric functional approach. *Comput Stat.* 2003;44:161–73.
32. Ferraty F, Vieu P. *Nonparametric functional data analysis: theory and practice.* New York: Springer-Verlag; 2006.
33. Locantore N, Marron JS, Simpson DG, Tripoli N, Zhang JT, Cohen KL. Robust principal component analysis for functional data (with discussion). *Test.* 1999;8:1–74.
34. Naya S, Francisco-Fernández M, Tarrío-Saavedra J, López-Beceiro J, Artiaga R. Application of functional data analysis to material science. In: Ferraty F, editor. *Recent advances in functional data analysis and related topics.* Berlin Heidelberg: Springer-Verlag; 2011.
35. Ramsay JO, Silverman BW. *Functional data analysis.* New York: Springer-Verlag; 2005.
36. Tarrío-Saavedra J, Naya S, Francisco-Fernández M, López-Beceiro J, Artiaga R. Application of functional ANOVA to the study of thermal stability of micro-nano silica epoxy composites. *Chemometrics Intell Lab Syst.* 2011;105:114–24.
37. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2011. <http://www.R-project.org>
38. Febrero M, Oviedo M. *fda.usc: The functional package: summary information.* R package version 1.0; 2009.
39. Febrero-Bande M, González-Manteiga W. Generalized additive models for functional data. *Test.* 2013;22:278–92.
40. Wood S. *Generalized additive models: an introduction with R.* Boca Raton: Chapman and Hall/CRC; 2006.
41. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen.* 1936;7:179–88.
42. Hill T, Lewicki P. *Statistics methods and applications.* Tulsa: StatSoft; 2007.
43. Fix E, Hodges JL. Discriminatory analysis, nonparametric discrimination: consistency properties. USAF School of Aviation Medicine, Randolph Field, Texas Technical Report 4; 1951.
44. Vapnik V. *Statistical learning theory.* New York: Wiley; 1998.
45. Ripley BD. *Pattern recognition and neural networks.* Cambridge: Cambridge University Press; 1996.
46. 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.
47. Wehrens R. *Chemometrics with R. multivariate data analysis in the natural sciences and life sciences.* New York: Springer-Verlag; 2011.
48. Wand MP, Jones MC. *Kernel smoothing.* London: Chapman and Hall; 1995.
49. Ruppert D, Sheather SJ, Wand M. An effective bandwidth selector for local least squares regression. *J Am Stat Assoc.* 1995;90:1257–70.