Statistical functional approach for interlaboratory studies with thermal data

Article *in* Journal of Thermal Analysis and Calorimetry · November 2014

DOI: 10.1007/s10973-014-4039-1

6 authors, including:

Salvador Naya
University of A Coruña
97 PUBLICATIONS **877** CITATIONS
SEE PROFILE

J. Tarrío-Saavedra
University of A Coruña
76 PUBLICATIONS **683** CITATIONS
SEE PROFILE

Jorge López Beceiro
University of A Coruña
61 PUBLICATIONS **1,000** CITATIONS
SEE PROFILE

Mario Francisco-Fernández
University of A Coruña
73 PUBLICATIONS **562** CITATIONS
SEE PROFILE

J Therm Anal Calorim (2014) 118:1229–1243
DOI 10.1007/s10973-014-4039-1

Statistical functional approach for interlaboratory studies with thermal data

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

Received: 30 November 2013 / Accepted: 16 July 2014 / Published online: 26 August 2014

Abstract A new statistical functional data analysis (FDA) approach to perform interlaboratory tests is proposed and successfully applied to thermogravimetry (TG) and differential scanning calorimetry (DSC). This functional approach prevents the typical losses of information associated to the dimension reduction processes. It allows the location and variability of the thermal curves obtained by the application of a particular test procedure. The intra- and inter-laboratory variability and location have been estimated using a FDA approach as well as the traditional reproducibility and repeatability studies. To evaluate the new approach, 105 TG curves and 90 calorimetric curves were obtained from calcium oxalate monohydrate. The obtained curves correspond to seven simulated laboratories, 15 curves per laboratory. Functional mean and variance were estimated. From a functional point of view, these descriptive statistics consider each datum as a curve or function of infinite dimension. Confidence bands were computed using smooth bootstrap resampling. A laboratory consistency study is performed in a functional context. The functional depth approach based on bootstrap resampling is a useful tool to identify outliers among the laboratories. The new FDA approach permits to identify as outliers the thermal curves obtained with old or wrong calibrations. Functional analysis of variance test based on random projections and the false discovery rate procedure (FDR) provides which laboratories obtain significant different thermal curves. This approach can be applied to perform interlaboratory test programs where the response of the test result is functional, as, for example, DSC and TG tests, without having to assume that data follow a Gaussian distribution.

Keywords Interlaboratory study · Functional data analysis · Thermogravimetry · Differential scanning calorimetry · Laboratory consistency · Data depth

S. Naya (✉) · J. Tarrío-Saavedra · J. López-Beceiro · R. Artiaga
Escola Politécnica Superior, Universidade da Coruña, Ferrol, Spain
e-mail: salva@udc.es

M. Francisco-Fernández
Facultade de Informática, Universidade da Coruña, A Coruña, Spain

M. Flores
Escuela de Ciencias Físicas y Matemáticas, Universidad de Las Américas, Quito, Ecuador

Introduction

Nowadays, advances in computer technology, the continuing growth of computing power and data storage capacity, have made some techniques of statistical data analysis, before unapproachable, competitive to traditional univariate and multivariate analysis techniques. This is the case of functional data analysis (FDA). FDA techniques tackle each function or curve of infinite dimensions such as a datum, preventing the information loss derived from dimension reduction procedures as principal component analysis or the use of a point of the curve. Therefore, these techniques could be very useful to deal with thermal data, including thermogravimetry (TG) and differential scanning calorimetric (DSC) curves in contexts such as the regression, classification or analysis of variance (ANOVA) problems [1–4].

Nevertheless, the use of FDA techniques is still uncommon in statistical quality control tasks applied to analytical chemistry, as it is the case of the interlaboratory

studies (ILT). In the present study, a new approach to perform ILT combined with the application of FDA data processing tools is proposed. In this work, the proposed methodology is applied to TG and DSC curves.

Interlaboratory studies

An ILT can be defined as the statistical quality control process used to evaluate the consistency of the test results obtained using a well-defined experimental process, applied to the same specific controlled material and tested by several different laboratories [5, 6]. It is important to note that the ILT test results corresponding to a specific test procedure can be used to evaluate the precision for that test [5, 6].

The aim of the present study is to introduce and explain a statistical experimental design based on the use of thermal analysis techniques and FDA data processing. We seek to construct an ILT to estimate the precision of the results obtained by an analytical test procedure, as it is done in traditional repeatability and reproducibility studies (r and r) [5, 6]. In a traditional ILT, a test method or procedure is performed by different laboratories (more than six) on the same controlled material. We want to seek all of the potential variations that can occur when the test method is used [5, 6]. Many sources of variability have to be taken into account: different operators, different equipment, different environmental conditions, etc. Thus, we should expect to have greater variability among the results from different laboratories. The measure of this larger variation due to readings taken by different laboratories is found as the reproducibility variance [5, 6]. The repeatability variance is corresponding to the variations intralaboratories [5, 6]. There are many interesting works about the ILT studies [7, 8].

Statistical functional data analysis techniques

FDA represents a new way to handle data in thermal analysis. FDA could represent a more suitable way for thermal data analysis than the traditional univariate and multivariate statistical approaches in the special case of working with thermal curve characteristics.

A random variable X is called functional variable if it takes values in an infinite dimensional space [1]. Thus, an observation x of the variable X is called a functional data. The functional variables are usually associated with processes continuously monitored in time, considering that these variables follow a continuous functional relationship. TG and DSC data are functional data since they are continuous functions defined on an interval.

Functional descriptive analysis, including functional mean and variance, is described in [9]. The data depth concept explains how a datum is centered with respect to a set of points belonging to a given population. Therefore, the deepest datum will be that surrounded by the highest number of neighbors. In FDA context, deeper curves are identified as those closer to the center, which are usually estimated by the median [10]. Three of the most common approaches to calculate the functional depth are the depth of Fraiman and Muniz (or median depth) [11], the mode depth [10], and the depth based on random projections [12]. The use of functional data depth can be used for outlier detection. This is a very important issue in ILT.

Regarding the ANOVA studies, the target is to verify if some continuous response variable (for example, temperature, mass loss, modulus, etc.) is significantly different depending on the value of one or more factors, explanatory variables defined at some specific levels (e.g. quantity of silica added at three different levels, two different operators, four different heating rates, etc.). In the FDA context, functional ANOVA, a test for functional response (where each datum is a TG curve) and categorical explanatory variable (three different mass percent silica additions) was designed and developed in [13]. This test has been applied to check if the amount of silica fume nano- and microparticles significantly affects to the composites thermal stability, estimated by the TG curves [1]. Other approaches to the same problem can be found in Brumback and Rice [14], Ramsay and Silverman [9, 15], Ramsay and Hooker [16], Fan and Lin [17], Shen and Faraway [18]. Recently, a study related to random projection procedure that allows the use of more than one factor was proposed [19]. This last technique corresponds to the approach used in the present study.

In the present study, we use R statistical software [20], considered *lingua franca* of statistical research. The R statistical software is the most adequate option to perform FDA studies just because there are many tools and functions programed nowadays. Some of the principal tools packages are the `fda` and `fda.usc` [21]. They make easier and more productive the application of FDA statistical techniques.

Experimental

Calcium oxalate monohydrate (99.0 + % purity) by Panreac, $Ca(COO)_2{\cdot}H_2O$, is used in this simulated ILT program to validate the present thermal analysis test procedure. Calcium oxalate is a material commonly used in thermal analysis for checking purposes because it presents well defined and separate thermal transitions, at three different temperatures in a range between 40 and 800 °C.

Ninety calcium oxalate samples were tested in a SDT 2960 TA Instruments simultaneous analyzer by six different operators (fifteen by each one of them). The intention is

Table 1 Laboratories scheme for the ILS program

Laboratory 1	An STA device with an old calibration program is used
Laboratory 2	An SDT simultaneous analyzer is used. We suppose that they follow all the specifications defined by the ILT study.
Laboratory 3	
Laboratory 4	
Laboratory 5	
Laboratory 6	An SDT simultaneous analyzer is used with an old calibration.
Laboratory 7	An SDT simultaneous analyzer is used but applying a biased calibration (2 °C offset from the zinc melting point).

to simulate the operation of a core group of four laboratories providing results more similar to each other. A heating ramp of 20 °C min^{-1} was applied in the range from 20 to 900 °C, under an air flow rate of 50 mL min^{-1}. Fifteen samples of an operator were tested using a deliberated erroneous temperature calibration, applying a bias of 2 °C with respect to the proper zinc melting temperature. This was done to simulate the case where the measures performed by a specific laboratory are biased because of temperature calibration problems. Moreover, another fifteen samples were performed using ancient calibration information (2 years old), and of course obtained by other different operator. The remaining sixty samples corresponding to other four different operators were tested using the same calibration conditions. We supposed that these results were obtained following properly all the indications of this ILT program, thus, this core group of resulting TG and DSC curves could represent the most unbiased, accurate, and reliable set of the sample.

In addition, fifteen calcium oxalate samples were tested in a rheometric STA 1500 simultaneous analyzer to obtain TG curves corresponding to a laboratory that could use other type of device. An old calibration program was also used in this case. Therefore, a case of malpractice is also intentionally included.

Overall, 105 oxalate samples were tested, obtaining two different types of curves: DSC and TG curves. It is important to stress that we have obtained 105 TG curves and 90 DSC curves. The DSC curves obtained from the STA device were not used to study properly the variability using only the SDT device. Table 1 presents the laboratories, summarizing the characteristic experimental conditions for each case.

It is important to note that in this study, we use two different instruments (simultaneous analyzers) as mentioned above, and these instruments provide two types of data: DSC curves and TG curves (Table 1). Thus, the TG curves corresponding to both types of devices are analyzed. Otherwise, the DSC curves are analyzed using just one type of device (SDT). The DSC curves obtained by the STA instrument are previously not considered in this study because they are extremely different from the DSC curves obtained by the SDT device. In a first iteration, they were considered outliers and, therefore, not included in the present work.

FDA methodology applied to interlaboratory studies

The aim of the application of this new methodology is to perform an ILT when the experimental results are curves, in this case TG and DSC curves. The goal is to consider the information of the whole thermal curves, instead of extracting just one feature from them and applying a standard univariate ILT. In this way, the loss of information derived from feature extraction (for example, defining a DSC curve using only the information of the glass transition temperature) is prevented.

As pointed out in the introduction, we seek to estimate the mean response obtained by the test procedure, the overall variability, the variability corresponding to each laboratory TG–DSC curves, to statistically test if the mean response (TG–DSC curves) is significantly different depending on the laboratory and performing a new approach for detecting those laboratories providing results significantly different from the remaining. In fact, an alternative consistency study of the laboratories results is performed in a functional context. We seek to know the laboratories providing non-consistent TG or DSC curves.

It is important to note that this new methodology can be applied just in any ILT where the experimental results are curves. In order to encourage practitioners, in this section, all the statistical FDA techniques applied for these purposes and the corresponding computing procedures are briefly described.

The first step of statistical data analysis is to implement the descriptive analysis. Thus, a first idea about the similarities and differences between laboratories can be obtained estimating the functional mean and variance [9]. FDA techniques allow to estimate the functional mean and variance (they are functions, not vectors or univariate variables) using punctual sample estimators (they will be also curves) and confidence bands (they will be areas, similar to the regression confidence intervals) [9]. These tools can help to identify and discard those laboratories that provide results different from the others, and further estimating the actual location and variability of the resulting TG and DSC curves. The intralaboratory variability is estimated calculating the functional variance for each independent laboratory, while an idea about the inter-laboratory variability is obtained comparing the point wise TG of DSC means corresponding to each laboratory. For more information about the mean and variance estimation in a

functional way, consult the works of Ramsay and Silverman [9] and Cuevas et al. [12].

To perform an adequate laboratory consistency study, the implementation of outlier detection methods and ANOVA test is necessary. Therefore, in a functional context, the mode, median, and random projection data depth methods [11, 12, 19, 22] are used to identify the non-consistent laboratories (outliers). These methods actually obtain the TG and DSC curves detected as outliers; thus, the laboratories with a larger quantity of atypical thermal curves are susceptible to be outliers. The implementation of these procedures needs the use of proper computing tools. The recent development of computing science gives the possibility of applying these new FDA techniques. The procedure for outlier curve detection is defined in the following subsections.

Finally, for properly identifying the laboratories that provide non-consistent results, the implementation of a functional ANOVA test is pertinent. This provides numerical results about which laboratories are significantly different. Combined with atypical curve detection, using functional data depth, the functional ANOVA test identifies the outlier or non-consistent laboratories: first of all, the outlier TG–DSC curve detection methods are implemented, and then the functional ANOVA test is applied to assess if the laboratories defined with more outlier curves are significantly different from the others. More information about this test can be obtained in Cuesta-Albertos and Febrero-Bande [19].

The calculations with functional data require adequate computational programs. The free statistical software R [20] was used to perform data analysis in a functional context. The well-known R software was mainly developed for statistical and graphical applications. It contains a complete set of libraries related to statistical tools, including regression, design of experiments, functional and special statistics, quality control, multivariate analysis, inference, data mining, etc. Thus, it was applied for estimating the functional mean and the variability related to the repeatability and reliability of a specific test procedure, to perform an ANOVA study, and to propose a new methodology for outlier detection, i. e., the laboratories providing biased and/or not accurate measures. The R packages `fda` [16] and `fda.usc` [21] were used for these purposes. Package `fda` is a basic reference to work in R with functional data. On the other hand, the package `fda.usc` [21] was applied for calculating functional statistics, outliers based on depth concept, and to perform functional ANOVA studies based on the random projection [19] concept.

Descriptive statistic calculation by bootstrap resampling

In this section, a bootstrap methodology for functional data is presented. Bootstrap [10] is a resampling procedure that permits the estimation of the sampling distribution of almost any statistic in a very simple way. It is a useful and more flexible alternative to the parametric methods (for example, assuming a Gaussian distribution) in statistics inference.

The aim is to obtain a descriptive idea about the location and variability of the TG and DSC curves, considering all the data together or making the statistical calculations for each laboratory. Thus, we can get an approximation of the overall functional mean and variance of the results obtained by the test procedure (without taking the risk of losing important information as could take place in the multivariate and univariate approaches). Moreover, the functional means and variances corresponding to each laboratory can be compared to identify anomalous differences. These tools can help us to identify and discard those laboratories that provide results different from the others. Then, the actual location and variability of test procedure results can be properly estimated.

In this study, we apply the smoothed bootstrap procedure as a resampling methodology for functional data. Using this method, the functional mean and variance statistics, with their bootstrap confidence bands can be obtained and represented. This procedure fills the holes in the functional space taking into account the covariance structure in the smoothing [10, 21].

Given the original TG curves (or alternatively the DSC ones), $X_1(t), ..., X_n(t)$, realizations of a functional random variable X of infinite dimension, defined by its functional mean and its functional variance, the bootstrap confidence bands are computed as explained below:

(a) *nb* bootstrap samples are calculated from the distribution of X: $X_1^{*(i)}(t), ..., X_n^{*(i)}(t)$, where $X_j^{*(i)}(t) = X_j(t) + Z(t)$, with $i = 1, 2, \ldots, b$ and $j = 1, \ldots, n$, $Z(t)$ is a normally distributed variable with mean equal to 0 and covariance matrix $\gamma \sum_x$, $\sum_x$ is the variance covariance matrix of the experimental sample composed of $X_1(t), ..., X_n(t)$, γ the smoothing parameter that controls the amount of variability of the new bootstrap samples.

(b) The required characteristic (i.e. functional mean, marginal variance, median, mode, etc.) is calculated from the real sample, $T(X_1(t), ..., X_n(t))$, and from each bootstrap resample of the TG or DSC curves, $T(X_1^{*(i)}(t), ..., X_n^{*(i)}(t))$, denoted by T^{*i}.

(c) Then, the distances between the T value obtained from the real sample and the T's ones obtained from each one of the *nb* bootstrap resamples are calculated using the L_2 norm, $d(T, T^{*i})$, with $i = 1, 2, \ldots, nb$. Afterward, a bootstrap confidence ball corresponding to a $1 - \alpha$ confidence level is defined such that the distance between T and some TG–DSC curve

corresponding to this ball was $d(T, X) < d_\alpha$, where d_α is the quantile $1 - \alpha$ of the L_2 distances between the actual sample statistic and the bootstrap resamples

It is important to note that the confidence bands give an idea of the uncertainty of the calculated statistic, namely mean and variance, since these statistics are actually random variables. We also have to stress that the confidence interval estimates in a functional context are not usual. The present work try to generalize them; in fact, in this work, the TG and DSC functional means are estimated using functional confidence bands (in the following marked as the colored area surrounding the point estimate).

The `fdata.bootstrap` function, included in the `fda.usc` library, is used to estimate the functional mean, median, mode, and marginal variance [10]. The defined parameters in the present study were $nb = 150$ bootstrap resamples, a degree of smoothing equal to $\gamma = 0.1$ (`smo` argument) and a signification level $\alpha = 0.05$, attending to the literature [21, 22]. The functions `func.mean` and `func.var` have to be passed as arguments in the function `fdata.bootstrap`.

Outlier detection using functional data depth

The statistical data depth notion has been proposed to detect outliers. In fact, it is noteworthy to observe that an outlier is defined by a low depth value. Following the definition of depth, an outlier is not surrounded by a large amount of points, thus, its depth tends to be lower than that of the other points. Inversely, the median of a dataset will have the higher depth value. For this reason, data depth analysis becomes a method for detecting outliers; they will be the points with lower depth values. In univariate data, the median is considered as the deepest point of a cloud of points. In a functional context, there are three main data depth measures based on the functional mode [12], median [11, 12], and random projections [12, 19, 22]. To obtain more information about the expression of these three types of functional depth and how to compute them, see [12].

The outlier detection could be crucial for obtaining a more objective criterion to identify and discard those laboratories that differ generally from the others. Of course, this is a very useful technique to identify and delete those anomalous results that can be occasionally obtained by the laboratories, to better define the location and variability of the final results without removing any laboratory from the ILT study.

In the present work, a procedure for obtaining outliers based on weighting combined with bootstrap resampling is used [22]. The function `outliers.depth.pond`, included in the `fda.usc` package, is used for this purpose.

According to this method, a way to detect the presence of functional outliers can be done observing the curves with lower depths:

(a) The functional depths corresponding to the experimental TG or DSC curves, $X_1(t), \ldots X_n, (t)$, are calculated applying the Fraiman Muniz, Mode or Random Projections depths $D_n(X_1(t)), \ldots, D_n(X_n(t))$.
(b) Then, $B = 100$ standard bootstrap resamples $X_1^b(t), ..., X_n^b(t)$ are obtained from the original TG-DSC curves, with $b = 1, \ldots, B$. Each resample is obtained with a probability proportional to its depth.
(c) New smoothed bootstrap are obtained $Y_i^b(t) = X_i^b(t) + Z_i^b(t)$, where $i = 1, .., n$ and $Z_i^b(t_1), ..., Z_i^b(t_m)$ (with $t_1, \ldots, t_m$ the time, temperature, etc. where the curves are defined), follow a Gaussian distribution with mean 0 and covariance matrix $\gamma\Sigma_x$, obtained from the covariance matrix of $X_1(t), ..., X_n(t)$, and using the bootstrap smoothing parameter γ.
(d) A C^b value is obtained from each bootstrap set $b = 1, \ldots, B$ corresponding to the empirical cth percentile of the depths distribution, $D_n(Y_i^b(t))$, with $i = 1, \ldots, n$.
(e) We obtain C as the median of the B different C^b values.
(f) The TG or DSC curves such that $D_n(X_i) \leq C$, for the obtained cut-off C, will be considered outliers and will be deleted then from the sample.

The following parameter values were chosen in order to perform the procedure using R software: $B = 100$, $\gamma = 0.05$, and $\alpha = 0.1$ (cth = 10 %), attending to the recommended ranges in the literature [21, 22].

Functional analysis of variance using random projections

The ANOVA analysis provides information about whether there are significant differences between the mean responses corresponding to each level of the factor. This procedure provides an objective criterion for discerning which levels produce different responses. In the present study, the response is functional, the TG and DSC curves, and the factor is a qualitative variable composed of 7 (for TG curves) or 6 (for DSC curves) levels or laboratories. Using this structure, we can implement a Functional ANOVA test. This approach offers the advantage of preventing the loss of information that could occur when using the traditional univariate and multivariate approaches. Functional ANOVA test performs the following statistical contrast, where k is the number of levels and m_i is the functional mean of the level i:

H_0: $m_1 = m_2 = \cdots = m_{k.}$
H_1: at least one of the means is different from the others.

We can perform Functional ANOVA analysis using the `anova.RPm` function provided by the `fda.usc` library.

The implemented procedure is based in the Random Projection procedure, consisting of the analysis of k randomly chosen one-dimensional projections [19]. This function performs ANOVA tests for functional response (in the present case, the TG and DSC curves), and continuous or not covariates (in this study, the different laboratories are qualitative covariates). This function also allows the application of special pairwise tests.

The main aim of the random projection functional ANOVA method is to obtain univariate random projections from functional data (TG–DSC curves), then, solving the univariate ANOVA problem for each projection, and finally to employ the information obtained from k projections to develop conclusions about the original functional data [19]. To perform the univariate ANOVA test obtaining reliable results, even under heteroscedastic and/or non-Gaussian conditions, the test proposed by Bruner et al. [23] was used in fact, it is recommended by Cuesta-Albertos and Febrero-Bande [19].

As we manage k random projections, it is necessary to apply a correction to the resulting p-values, in order to avoid the risk of rejecting the null hypothesis when this is actually true. The well-known Bonferroni's correction [24] could be too much conservative, and a bootstrap correction could be a time-consuming process. Therefore, Cuesta-Albertos and Febrero-Bande recommend to employ a correction designed to control the false discovery rate (FDR) [19], where the FDR is the expected proportion of wrongly rejected hypotheses, when k null hypothesis is tested, one per projection. Accordingly, the null hypothesis of equality of functional means will be rejected when $\alpha \geq \inf\left\{\frac{k}{i}p_{(i)}, i = 1, \ldots, k\right\}$, being the overall corrected p-value $= \inf\left\{\frac{k}{i}p_{(i)}, i = 1, \ldots, k\right\}$. The k value is the number of projections, p_1, …, p_k are the p-values corresponding to the evaluation of the null hypothesis for each projection, and $p_{(1)},\ldots,p_{(k)}$ the ordered sequence.

Following the indications of [19], `RP = 30` random projections were used in each case. In addition, pairwise comparisons were obtained using the same procedure. These were implemented applying `contr.sum` function that develops a matrix where the last factor level (laboratories) is compared to the remaining ones. This function was performed for seven (TG curves) and six (DSC curves) levels of the factor (laboratories). `contr.sum` (7) or `contr.sum` (6) has to be included in the `contrast` parameter of `anova.RPm`.

Results and discussion

The proposed ILT methodology has been tested by its application to real TG and DSC curves corresponding to controlled calcium oxalate samples. The results obtained by the test procedure described in the experimental section have been analyzed following a functional ILT approach.

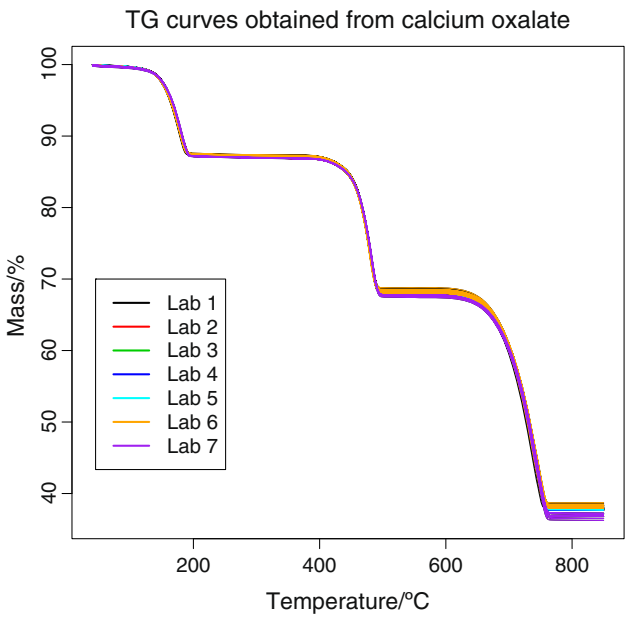

Fig. 1 Calcium oxalate TG curves obtained by the seven laboratories

Inter alia the objectives of ILT are the correct estimation of the process variability and calculating a position value that defines the result set: mean, median, or mode. As the TG and DSC curves can be treated as functional data [1–4], in this section, we have presented the application of the FDA tools to estimate their position and variability measures and also to propose an alternative for laboratory consistency studies in a functional context. At the end of this section, the univariate ILT methodology defined by the ASTM E691 [5] is applied in order to compare its results with the present functional approach.

Thermogravimetric curves study

Figure 1 shows the TG curves obtained for each laboratory. As expected, we can see three well-defined and separate degradation processes under oxidative conditions. The obtained TG curves appear to overlap, close to each other. It appears that laboratory 7 provides TG curves with a small bias with respect to the others in the last step.

The estimation of the functional mean and variance is necessary to get more information about the location and variability results. Then, overall functional mean and variance are calculated [9, 15, 25]. The results are plotted in Fig. 2. In addition, the confidence bands for the mean TG curves estimation are calculated using a confidence level of 95 %, i.e., the functional version of the confidence intervals to estimate the actual TG mean (the population mean).

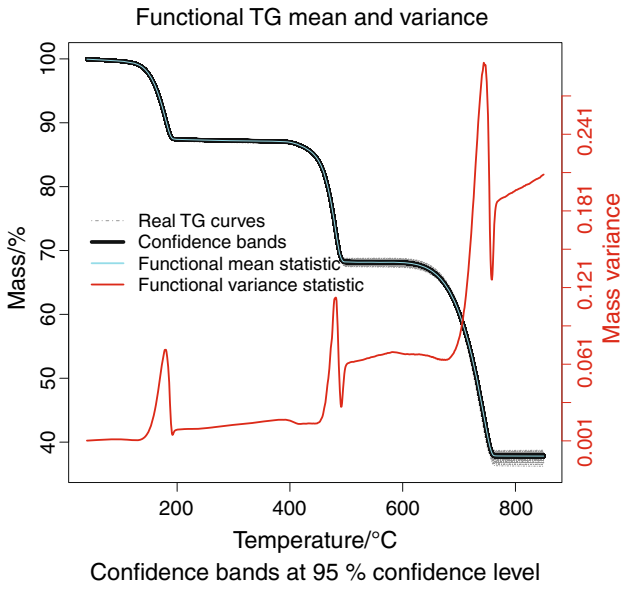

Fig. 2 TG curves, functional mean estimation, and its confidence bands with 95 %. Functional variance estimation corresponding to the calcium oxalate TG curves

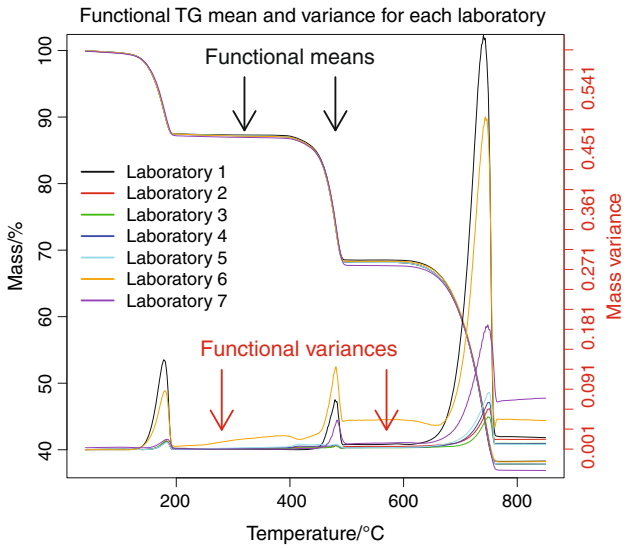

Fig. 3 Functional mean and functional variance estimations corresponding to the calcium oxalate TG curves obtained by each laboratory

As we can observe in Fig. 2, the TG mean estimate for calcium oxalate and the proposed test procedure presents a very narrow confidence bands. This is an indicative of a low variability in the mean estimation (high precision of the functional mean estimate) and, thus, the goodness of the estimate.

Focusing on the functional variance estimate of the TG curves, a continuous increasing variability, depending on temperature, is shown in Fig. 2. The level of variance becomes extreme in the last step of degradation, when compared with the functional mean values of TG curves at that temperature. We can infer that the variability of the measured process is higher at higher temperatures. Even, we could suggest that the bias from the functional mean could be accumulative with respect to the increasing temperature for the special case of the TG curves. To assess this, the same mean and variance calculation should be done for each laboratory. Moreover, the traces of the functional variance show peaks of higher values at the temperatures, where the maximum rate of degradation is reached for each degradation process. The uncertainty is greater at these temperatures, where the control of programmed experimental conditions becomes more difficult.

The functional means and variances have also calculated to measure the location and variability of the test procedure results depending on the chosen laboratory. Figure 3 shows the results. Focusing on functional TG means for each laboratory, all the traces are very similar, they are close to each other, although the functional mean of the TG curves corresponding to laboratory 7 slightly differs from the others in the last process of degradation. If the functional variances are observed, then the continuous increasing of overall variance shown in Fig. 2 is largely due to the high and also increasing functional variance of the TG curves obtained by laboratory 6. In fact, laboratory 6 used a wrong temperature calibration (Table 1). Then, this result is reasonable. Also, abnormally high functional variances have been observed for laboratories 1 and 6, when they are compared with those corresponding to the remaining ones. These differences are obtained mainly at the variance peaks. These variance peaks occur at the temperatures where the maximum rate of degradation is reached for each degradation process. The functional variances seem to be very small in the remaining laboratories. Still, an increase in the variance peak heights is especially observed in the last step of degradation. This trend is even greater if the variance peaks are compared with the values of the corresponding functional mean at each temperature.

As a result of the random projections and mode depth calculation for each curve shown in Table 2, we conclude that the TG results obtained by laboratories 1, 6, and 7 are possible outliers, and then they are susceptible to be deleted from the final ILT estimations. All these results are reasonable since these laboratories used wrong calibration information (laboratories 1, 6, and 7) and a different test instrument (laboratory 1), see Table 2. Moreover, it seems that the case that provides a bigger bias in TG curves is the wrong temperature calibration (see Fig. 2 and 3), while the variability is mainly affected by the odd calibration. The use of a different instrument, a STA device, could increase the variability, at least in this case (see the functional variance corresponding to the laboratory 1 in Fig. 3).

Table 2 Random projection depths corresponding to the different TG curves obtained from calcium oxalate samples

Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
0.1351	0.2170	0.2540	0.1586	0.2445	0.1212	0.1232
0.0795	0.2282	0.1785	0.2247	0.2084	0.0914	0.1376
0.1361	0.2010	0.2240	0.2401	0.1816	0.1750	0.1135
0.1251	0.2159	0.2330	0.2159	0.2190	0.0995	0.1010
0.1022	0.2361	0.2207	0.2073	0.1770	0.1953	0.1269
0.1188	0.1988	0.2113	0.2204	0.2330	0.1252	0.1334
0.1251	0.1785	0.1900	0.2240	0.1543	0.1212	0.1376
0.1437	0.1769	0.2251	0.2396	0.2220	**0.0698**	0.1144
0.1188	0.2127	0.1899	0.2247	0.1945	0.1134	0.1169
0.1188	0.1711	0.2464	0.2401	0.2282	0.1351	0.1121
0.1033	0.2309	0.2540	0.2328	0.2127	**0.0879**	**0.0827**
0.1361	0.2401	0.2190	0.1597	0.1982	0.1498	0.1121
0.1331	0.2404	0.2251	0.2309	0.2309	0.1631	**0.0584**
0.0644	0.2240	0.2129	0.2363	0.1721	0.1789	0.1076
0.0996	0.1945	0.2257	0.2404	0.2464	0.1578	**0.0793**

The TG curves corresponding to depths marked in bold are defined as outliers (TG depth < 10th percentile of the depths distribution, 0.0908)

In the present work, after the TG curves location and variability estimation, the second step of this functional descriptive analysis is to implement an outlier detection study. This is done to obtain more objective information about which laboratories could provide different results. Figure 4 shows the TG curves marked as outliers using the mode, median, and random projection depths [12], and the bootstrap procedure defined in the section named FDA methodology applied to ILT studies. Depending on the type of depth used, the number of TG curves marked as outliers differs. In fact, Tables 2 and 3 show the depths corresponding to the different TG curves, calculated using the mode and random projection methods, respectively. These methods represent the extreme results: calculating the mode depths we obtain the highest number of outliers, while if the random projection depths are measured, we reach the lowest number of atypical TG curves. In all the cases, $B = 100$ bootstrap resamples were chosen, and a smooth parameter $\gamma = 0.05$ was defined, using also a significance level equal to $\alpha = 0.1$ (cth is the 10 % percentile).

When the Random Projection depth is calculated, we obtain TG curves considered as outliers only for the laboratories 1, 6 and 7. Thus, we can guess that these laboratories provide TG curves different to those provided by the remaining laboratories. We obtain a bigger evidence when the bootstrap outlier detection procedure is applied using the calculated mode depths: all the TG curves corresponding to the laboratories 1 and 2 are marked as outliers, and almost all the TG curves obtained by the laboratory 6 are also detected as atypical data. Therefore, these are objective evidences that laboratories 1, 6 and 7 provide different results.

As pointed above, the Functional ANOVA analysis can provide objective information about whether there are significant differences between the mean responses corresponding to each level of the factor. This procedure represents the definitive criterion to discern which levels produce different responses. This is not a descriptive analysis tool, but an inference process.

Table 4 presents the adjusted p-values resulting from the different pairwise comparisons between the different levels. The p-values are obtained from the Bruner et al. test [23] and adjusted by the FDR procedure [19], as previously described. Using a standard confidence level equal to 0.95, p-values <0.05 indicate that the TG functional means corresponding to the two different laboratories compared are different with statistical significance. We can observe that there are not differences in the TG curves provided by the laboratories 2–5. These are the simulated laboratories using the same test instrument and calibration procedure. As we suppose that these laboratories properly followed the indications of the ILT program, the location and variability of the proposed test procedure results must be estimated using only the TG curves obtained by them. Otherwise, the results obtained by the laboratories 1, 6, and 7 that used old or wrong calibrations are statistically different from each other and from the ones obtained by the laboratories 2–5.

Summarizing, attending to the resulting p-values of the functional ANOVA test, we should delete the observations obtained by laboratories 1, 6, and 7. Then, as conclusion of the ILT program, we would be able to properly estimate the functional mean and variance of the results obtained by this test procedure.

Calorimetric curves study

The same functional ILT data processing applied to the TG curves is now applied to the obtained DSC curves. In this case, the ILT is performed using six laboratories: laboratories 2–7.

Figure 5 shows the DSC curves obtained for each laboratory. Again, the three reactions appear well defined and separated. As expected under air atmosphere, the first and third ones are endothermic, while the second one is exothermic. The DSC curves are close to each other, except the DSC curves provided by the laboratory 6 and some curves corresponding to the 7.

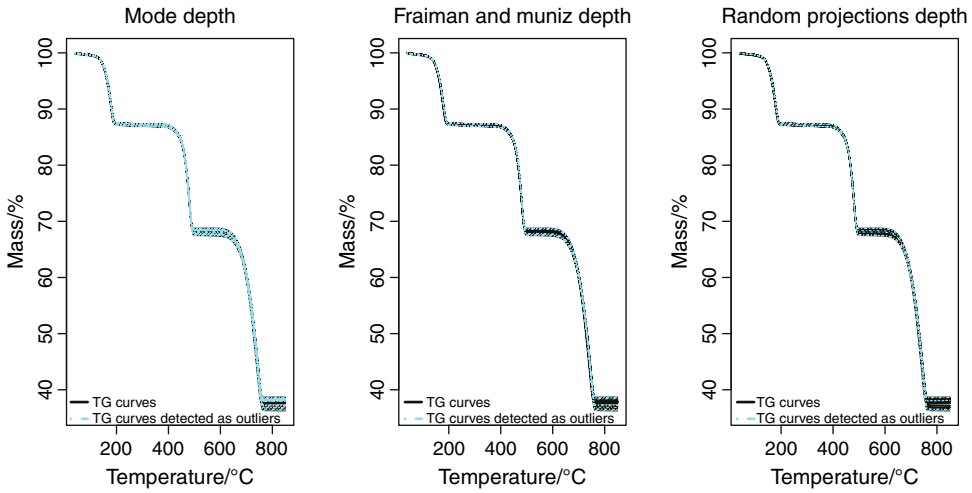

Fig. 4 Outlier detection applying functional depth procedures: mode, median (Fraiman and Muniz depth), and random projections depths

Table 3 Mode depths corresponding to the different TG curves obtained from calcium oxalate samples

Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
2.6027	11.9933	16.1837	6.6202	15.6488	**1.4947**	**1.6548**
1.1937	13.5966	8.4040	14.9200	13.4324	**1.2635**	**2.1337**
2.3395	13.1208	13.2219	15.3169	8.7713	4.4644	**2.3310**
1.6619	10.7575	14.4891	10.7575	13.4539	**0.7713**	**2.0267**
1.1812	14.6270	11.5517	12.6996	6.9916	5.9693	**1.9573**
1.7366	10.0207	11.9340	11.8748	14.4891	**0.8459**	**2.0380**
1.6619	8.4040	10.9403	14.0598	6.5118	**1.4947**	**2.1337**
2.4603	8.3704	14.1373	16.0011	13.1552	**0.8514**	**2.1775**
1.7366	11.9572	10.1641	14.9200	9.9451	**0.7906**	**1.3564**
1.7366	8.8463	14.7410	15.3169	13.5966	**2.6027**	**2.4958**
1.5871	13.1214	16.1837	15.3309	11.9572	**0.7170**	**1.2343**
2.3395	15.3169	13.4539	6.7424	12.3862	**2.0442**	**2.4958**
1.6871	14.5902	14.1373	13.1214	13.1214	3.8299	**0.5465**
0.4296	14.0598	12.4551	14.7803	7.7295	3.6674	**2.3760**
1.4124	9.9451	12.5093	14.5902	14.7410	4.1367	**1.0056**

The TG curves corresponding to depths marked in bold are defined as outliers (TG depth $<$ 10th percentile of the depths distribution, 2.9245)

Table 4 Resulting adjusted p-values obtained by the FDR procedure for pairwise comparison between the different levels of the factor (laboratory), using 30 random projections in a functional context

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
Laboratory 1	–	0.000	0.000	0.000	0.000	0.000	0.000
Laboratory 2	0.000	–	0.995	0.965	0.983	0.000	0.000
Laboratory 3	0.000	0.996	–	0.734	0.331	0.000	0.000
Laboratory 4	0.000	0.979	0.852	–	0.767	0.000	0.000
Laboratory 5	0.000	0.980	0.802	0.918	–	0.000	–
Laboratory 6	0.000	0.000	0.000	0.000	0.000	–	0.000
Laboratory 7	0.000	0.000	0.000	0.000	0.000	0.000	–

When p-value < 0.05, we suppose that the functional TG curves means of each pair of levels are actually different

As we can observe in Fig. 6, the DSC functional mean for calcium oxalate presents relatively narrow bootstrap confidence bands (see the subsection named Descriptive statistic calculation by bootstrap resampling) but wider than that calculated for the functional TG mean. We can infer that the DSC curves obtained by the SDT simultaneous analyzer present more variability than the TG ones. The DSC mean estimation is less precise than obtained from the TG curves.

Focusing on the functional DSC curves variance estimate, an increase of the variability is also observed at very high temperatures.

The functional DSC means and variances have also been calculated to measure the location and variability of the test procedure results for each laboratory (Fig. 7). Focusing on the functional DSC means for each laboratory, all the traces are very similar, except those corresponding to laboratory 6, which uses an old calibration. The functional variances present three positive peaks corresponding to the maximum and minimum of the DSC curve peaks. The variability is higher at temperatures where the rate of change of the oxidative reactions is also higher. The functional variances of each laboratory are more similar than in the case of TG curves. Nevertheless, it seems that the DSC curves obtained by laboratory 6 tend to present higher variance than the others.

So, we can guess that DTG results obtained by laboratory 6 are possible outliers and, then, it is susceptible to be deleted from the final ILT estimations.

The second step of this functional descriptive analysis is to implement an outlier detection study. Figure 8 shows the DSC curves detected as outliers using the mode, median, and random projection depths [12], and the

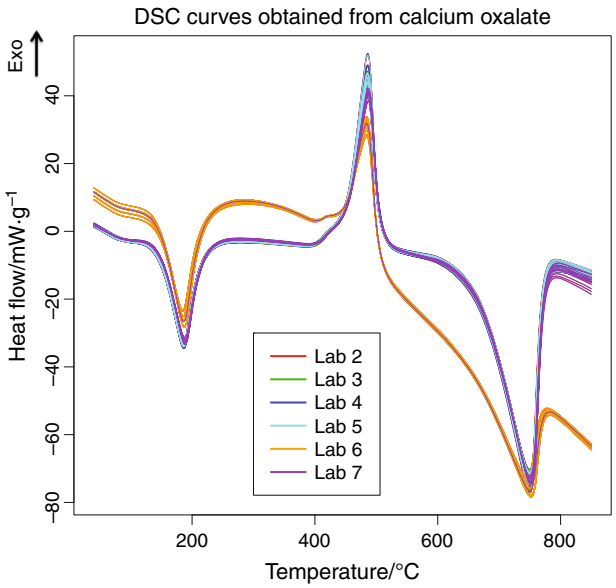

Fig. 5 Calcium oxalate DSC curves obtained by the six laboratories

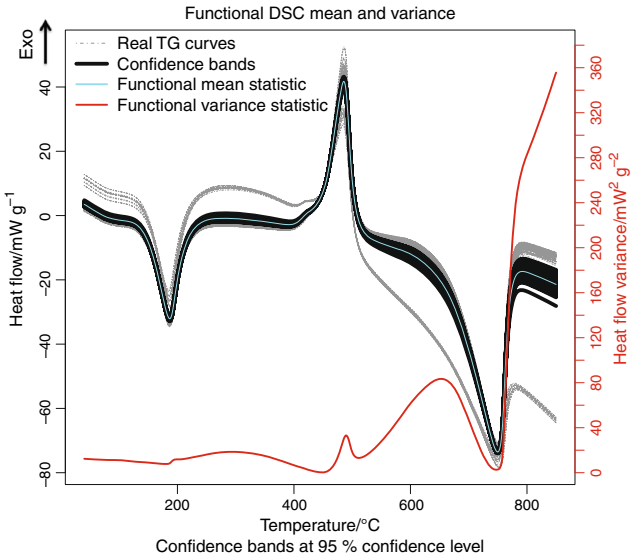

Fig. 6 DSC curves, functional mean estimation with its confidence bands and 95 %, and functional variance estimation corresponding to the calcium oxalate DSC curves

bootstrap procedure defined in the section named FDA methodology applied to ILT studies. Tables 5 and 6 show the depths corresponding to the different DSC curves calculated using the mode and random projection methods, respectively.

When the random projection depth is calculated, only three DSC curves corresponding to laboratory 6 are detected as outliers. Thus, this laboratory could provide DSC curves different to those provided by the remaining

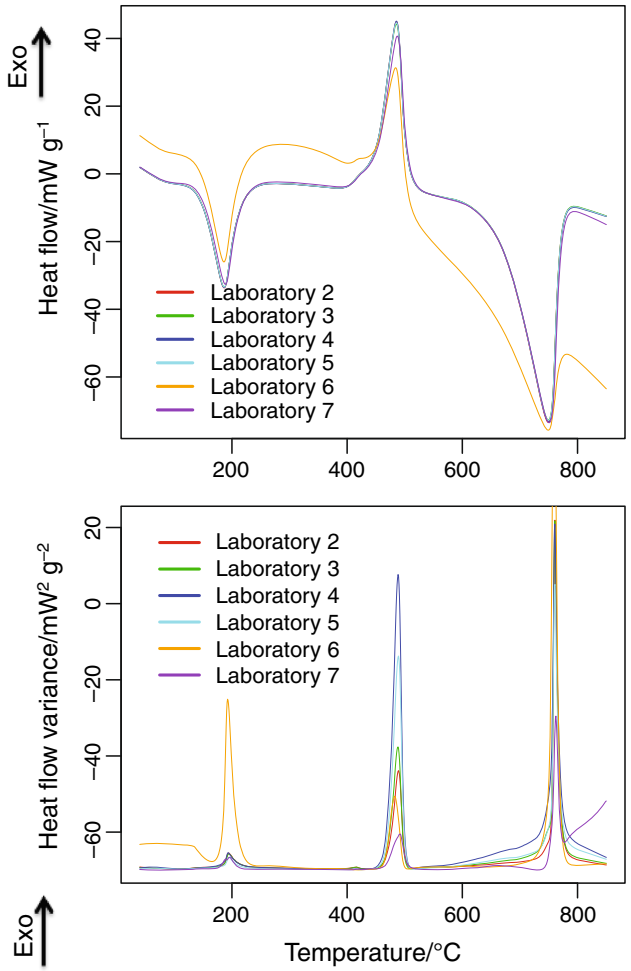

Fig. 7 Functional mean and functional variance estimations corresponding to the calcium oxalate DSC curves obtained by each laboratory

laboratories. We also obtain higher evidences, when the mode depths are calculated. All the DSC curves corresponding to laboratory 6 are marked as outliers, while three outliers were detected in laboratory 7.

Table 7 presents the adjusted p-values resulting from the different pairwise comparisons between the DSC curves of the different levels. We can observe that there are not differences between the DSC curves provided by the laboratories 2–5 (as in the case of TG curves). Otherwise, the results obtained by the laboratories 6 and 7 that used old and wrong calibrations, respectively, are statistically different from each other and from the ones obtained by laboratories 2–5.

Regarding the p-values of the functional ANOVA test applied to the DSC curves, we should delete the observations obtained by laboratories 6 and 7 to accurately estimate the mean and variance of the results.

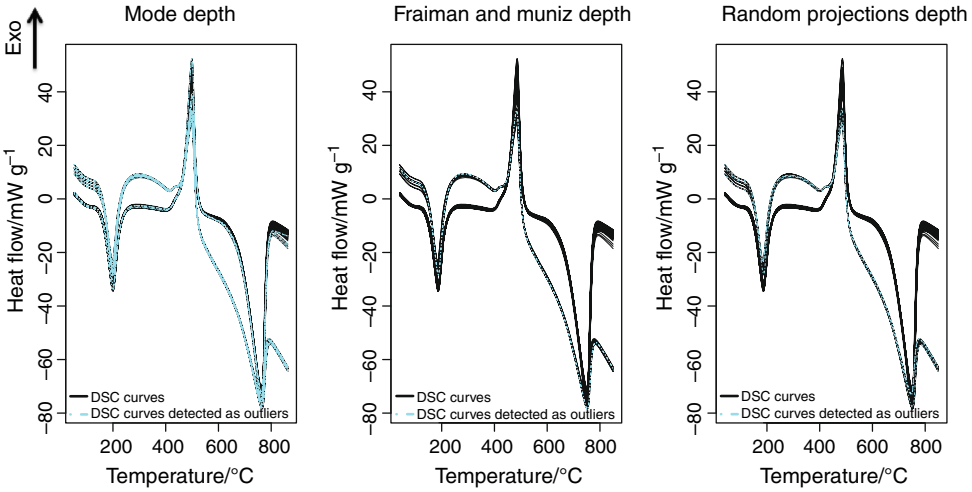

Fig. 8 Outlier detection applying functional depth procedures to DSC curves: Mode, Fraiman and Muniz, and random projections depths

Table 5 Random projections depths corresponding to the different DSC curves obtained from calcium oxalate samples

Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
0.1270	0.1775	0.2131	0.1400	0.0819	0.1976
0.1984	0.2291	0.1484	0.1963	0.0831	0.2030
0.2143	0.2319	0.2016	0.1746	**0.0744**	0.2016
0.2030	0.1433	0.2110	0.1824	0.1015	0.1666
0.1824	0.2202	0.2372	0.2110	**0.0797**	0.2041
0.1558	0.2016	0.2143	0.1334	0.0798	0.2096
0.2218	0.1830	0.2202	0.2291	0.0897	0.2016
0.2202	0.2112	0.1307	0.2372	0.1015	0.1831
0.2372	0.1973	0.1437	0.1882	0.0876	0.1686
0.2131	0.1848	0.1746	0.1973	0.0960	0.2041
0.1852	0.2131	0.1778	0.2233	**0.0703**	0.2032
0.2131	0.2246	0.1766	0.1963	0.0897	0.2016
0.2319	0.1717	0.2319	0.2157	0.1015	0.1746
0.1814	0.2065	0.1568	0.2325	0.1015	0.1844
0.1848	0.1433	0.2291	0.2016	0.0960	0.1779

The DSC curves corresponding to depths marked in bold are defined as outliers (DSC depth < 10th percentile of the depths distribution, 0.0798)

Table 6 Mode depths corresponding to the different DSC curves obtained from calcium oxalate samples

Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
6.1005	8.3484	10.6621	7.9054	**1.6322**	4.8224
10.8817	12.3794	4.9814	11.7619	**2.5315**	4.9910
12.4427	12.3152	12.1633	9.5054	**1.4780**	4.3486
11.0360	7.7439	12.3148	9.4926	**2.8170**	**1.4980**
9.4926	12.1509	13.8290	12.3148	**1.4432**	4.7700
9.1415	12.1633	12.4427	**2.5768**	**2.1526**	5.4513
12.9456	11.3906	12.1509	12.3794	**1.8701**	4.3486
12.1509	12.4332	**1.8293**	13.8290	**2.8170**	**3.1774**
13.8290	9.9901	4.5184	9.6577	**1.8610**	**1.8310**
10.6621	7.5364	9.5054	9.9901	**2.7553**	4.7700
10.7022	10.6621	9.7809	13.0506	**1.5444**	4.9724
10.6621	12.8747	8.0092	10.6083	**1.8701**	4.3486
12.3152	9.1126	12.3152	12.3090	**2.8170**	3.7312
8.8647	12.2934	9.1773	14.1676	**2.8170**	3.9567
9.9893	7.7439	12.3794	12.1633	**2.7553**	4.1255

The DSC curves corresponding to depths marked in bold are defined as outliers (DSC depth < 10th percentile of the depths distribution. 3.199)

Traditional univariate approach

In this subsection, the traditional methodology established by the ASTM E691 [5] is applied to compare the corresponding results with those obtained with the new functional approach above mentioned. A univariate response variable is needed. For this purpose, a specific feature related to thermal stability is extracted from the previously obtained TG curves just for comparing the univariate and functional methodologies. First, we used the initial decomposition temperature (IDT) [26], defined as the temperature at which the 5 mass% of the overall mass is

Table 7 Resulting adjusted *p*-values obtained by the FDR procedure for pairwise comparison between the different levels of the factor (laboratory), using 30 random projections in a functional context

	Lab 2	Lab 3	Lab 4	Lab 5	Lab6	Lab 7
Laboratory 2	–	0.908	0.505	0.940	0.000	0.000
Laboratory 3	0.853	–	0.665	0.997	0.000	0.000
Laboratory 4	0.400	0.631	–	0.856	0.000	0.000
Laboratory 5	0.993	0.949	0.819	–	0.000	0.000
Laboratory 6	0.000	0.000	0.000	0.000	–	0.000
Laboratory 7	0.000	0.000	0.000	0.000	0.000	–

When *p*-value <0.05, we suppose that the functional DSC curves means of each pair of levels are actually different

Table 8 IDT mean and standard deviation of the seven studied laboratories

	IDT mean	IDT standard deviation	Number of replicates
Lab 1	164.4	1.099	15
Lab 2	164.1	0.209	15
Lab 3	164.3	0.371	15
Lab 4	164.2	0.285	15
Lab 5	164.2	0.285	15
Lab 6	164.2	0.979	15
Lab 7	165.7	0.371	15

lost. In addition, another important feature of the TG is extracted and also studied: the mass loss in the second step of the TG traces. It is important to stress that the functional methodology provides mean, variability estimates, and outlier detection using all the information present in the whole TG curve, while the univariate procedure also provides mean, variance, and outliers based on just a small part of the whole TG data. This fact is the main difference between both approaches. Thus, our functional contribution will be applicable and useful as long as the laboratories results are curves.

Focusing on the oxalate IDT study, Table 8 shows the IDT mean and standard deviation corresponding to each laboratory.

In order to identify the outliers or laboratories that provide result significantly different from the others, the ASTM E691 proposes the use of h and k Mandel statistics [5]. The h and k statistics are computed to measure the consistence between the laboratory results. The h statistic measures the inter-laboratories consistency. It compares the replicate averages with the overall average: higher $|h|$ implies a less consistency than the remaining laboratories. On the other hand, the k statistic estimates the intralaboratory consistency. It compares the replicate standard deviations with the repeatability standard deviation. The h and k values of each laboratory must be compared with the critical limits h_{crit} and k_{crit}. The expressions for h, k, h_{crit} and k_{crit} are shown below:

$$h_{\text{i}} = \frac{\overline{x_{\text{i}}} - \overline{\overline{x}}}{\sqrt{\frac{1}{p-1}\sum_{\text{i}=1}^{\text{p}} \left(\overline{x_{\text{i}}} - \overline{\overline{x}}\right)^2}}; \quad h_{\text{crit}} = \frac{(p-1)\cdot \text{t}}{\sqrt{p\cdot t^2 + p - 2}},$$

$$k_{\text{i}} = \frac{S_{\text{i}}}{S_{\text{r}}} = \frac{\sqrt{\frac{1}{n-1}\sum_{\text{i}=1}^{\text{n}} (x_{\text{i}} - \bar{x}_{\text{i}})^2}}{\sqrt{\frac{1}{p}\sum_{\text{i}=1}^{\text{p}} S_{\text{i}}^2}}; \quad k_{\text{crit}} = \sqrt{\frac{p}{1 + \frac{p-1}{F}}},$$

where n, p, $\overline{x_{\text{i}}}$, and S_{i} are the number of replicates per laboratory, the number of laboratories, and the replicate means and standard deviations of each laboratory, respectively. Moreover, t is the quantile of a t student distribution with $n-1$ degrees of freedom and two-sided signification level equal to 0.005, in fact, $t_{\text{n}\,-\,1,\,0.005/2}$. Otherwise, F is the quantile of the F Snedecor distribution at a one-sided signification level equal to 0.005, and $(n-1)$ and $(n-1)\cdot(p-1)$ degrees of freedom, $F_{\text{n}\,-\,1,(\text{n}\,-\,1)\cdot(\text{p}\,-\,1),0.005}$. Finally, S_r is the repeatability standard deviation, while the reproducibility standard deviation (S_{R}) is the maximum value between the following two expressions:

$$S_{\text{R}}^{*} = \sqrt{\frac{1}{p-1}\sum_{\text{i}=1}^{\text{p}} \left(\overline{x_{\text{i}}} - \overline{\overline{x}}\right)^2 + S_{\text{r}}^2 \frac{n-1}{n}}; \; S_{\text{r}} = \sqrt{\frac{1}{p}\sum_{\text{i}=1}^{\text{p}} S_{\text{i}}^2}.$$

The h and k values for each laboratory are compared with the h_{crit} and k_{crit} limits. The laboratories defined with h or/and k values out of the limits (one sided for k statistic, two sided for h statistic) are considered outliers and removed. New k, h, h_{crit}, and k_{crit} values are then calculated until no outliers are obtained [5]. Table 9 shows the results of the previous expressions in the case of IDT variable study.

The h and k values calculated for each laboratory are shown in Figs. 9 and 10. The outliers are identified as the

Table 9 Statistical data analysis results corresponding to a standard ILT using the IDT feature and the ASTM 691 standard

Number of valid laboratories (p)	7
Overall mean ($\bar{x}$)	164.4
Repeatability standard deviation (S_{r})	0.6151
Reproducibility standard (S_{R})	0.9651
h_{crit}	2.05, −2.05
k_{crit}	1.44
Number of detected outliers	3
Laboratories defined as outliers	Lab 1, Lab 6, Lab 7

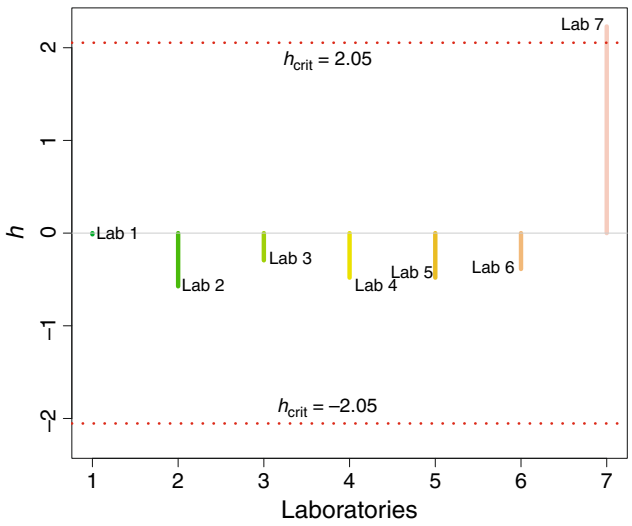

Fig. 9 Bar chart corresponding to the h statistic values of each laboratory obtained from IDT measures. Laboratory 7 (corresponding to the SDT device with a biased calibration) is identified as an outlier

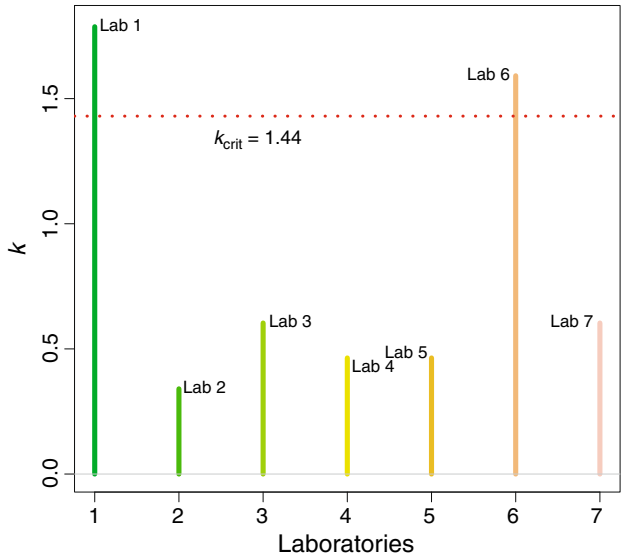

Fig. 10 Bar chart corresponding to the k statistic values of each laboratory obtained from IDT measures. Laboratories 1 (STA device with an odd calibration) and 6 (SDT device with and old and odd calibration) are identified as outliers

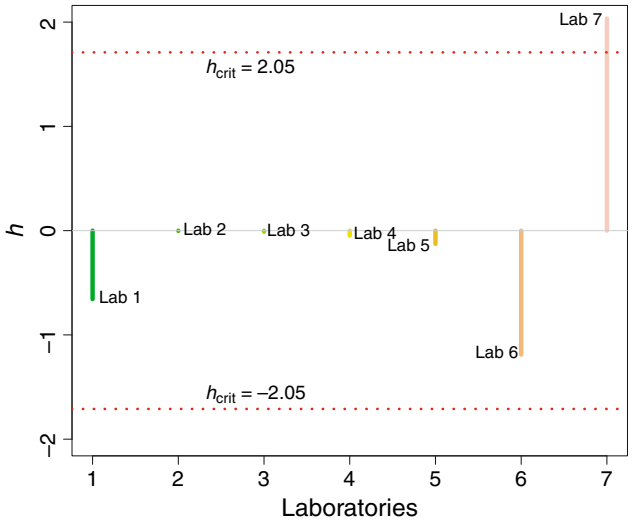

Fig. 11 Bar chart corresponding to the h statistic values of each laboratory obtained from TG second step mass loss measures. Laboratory 7 (corresponding to the SDT device with a biased calibration) is identified as an outlier

Table 10 Statistical data analysis results corresponding to a standard ILT using the variable TG mass loss in the second step and the ASTM 691 standard

	Iteration 1	Iteration 2
Number of valid laboratories (p)	7	6
Overall mean ($\bar{x}$)	30.36	30.28
Repeatability standard deviation (S_r)	0.0939	0.0613
Reproducibility standard (S_R)	0.4791	0.3327
h_{crit}	2.05, −2.05	1.92
k_{crit}	1.44	1.43
Number of detected outliers	1	0
Laboratories defined as outliers	Lab 7	–

values out of the limits defined by h_{crit} and k_{crit}. As in the case of the functional context, the laboratories 1, 6, and 7 were properly detected as outliers, they provide no consistent results with respect to the remaining. Therefore, both approaches, functional (taking all the curve information into account) and univariate (using just the information of one curve feature), provide the same result.

Otherwise, the feature or variable "mass loss in the second TG step" is analyzed using a standard ILT instead the IDT above mentioned. The mass loss in a TG single step is also an important value to characterize a degradation process. In this case, just laboratory 7 is identified as an outlier using the h and k statistics (see Table 10 and Figs. 11, 12), and in the second iteration, no outliers were obtained. Then, the wrong calibrated laboratories 1 and 6 are not detected. Therefore, we can conclude that the number of outliers identified using a

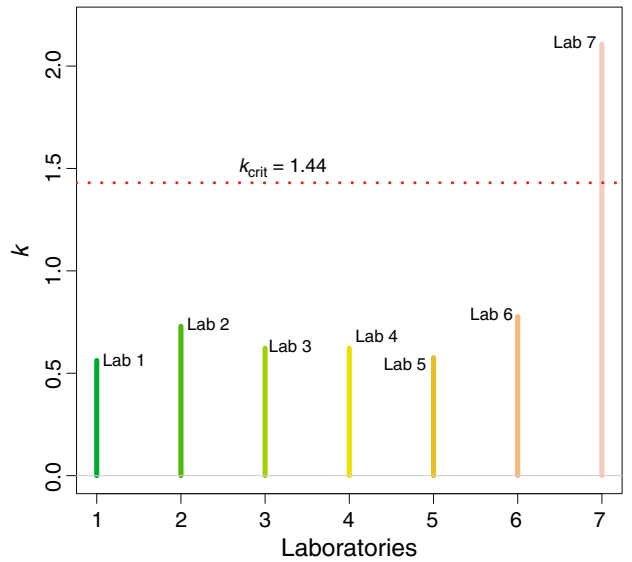

Fig. 12 Bar chart corresponding to the k statistic values of each laboratory obtained from TG second step mass loss measures. Laboratory 7 (corresponding to the SDT device with a biased calibration) is identified as an outlier

standard ILT, and also the same experimental curves, depends on the extracted feature (IDT, mass loss in the second step, etc.) used in the analysis. So, not all the features summarize properly all the information of a whole curve, The use of FDA techniques is justified because we can identify the non-consistent laboratories using the whole curves, before extracting the interest features in each thermal curve. Moreover, the functional methodology does not need

to fulfill any previous assumption as, for example, data normally distributed.

Conclusions

A new FDA methodology for ILT studies has been successfully applied to TG and DSC curves. It prevents from the typical losses of information associated to classical methodologies, and it is suitable for laboratory tests using functional response, like thermal analysis curves.

The usefulness of the proposed ILT methodology has been demonstrated by its application to real TG and DSC curves corresponding to controlled calcium oxalate samples. This approach estimates the location and variability of the functional thermal curves obtained by the application of a specific test procedure. Thus, this approach allows the validation of the test procedures.

The functional mean and variability intra- and inter-laboratories were estimated by functional descriptive analysis. The intralaboratories functional variability is higher at the temperatures where the maximum rate of change is reached for each reaction. The functional variance also increases at high temperatures. The functional statistic calculations give a first and rough idea about which laboratories could obtain different results from the others.

The functional depth approach is a useful tool for detecting the TG and DSC curves that could be outliers. All the "wrong-calibrated" laboratories were easily identified as outliers studying the whole TG curves. The TG curves are more useful than the DSC curves (simultaneously obtained using the same analyzer) for detecting atypical data.

In a functional context, the proposed approach allows us to detect significant deviations from the actual functional mean and variance of the resulting thermal curves. In fact, the laboratories that provide significant different results (thermal curves) from the others were successfully detected applying the functional ANOVA test based on random projections and the FDR procedure. As a result, there are no differences between the TG and DSC curves provided by the laboratories 2, 3, 4, and 5. However, the thermal curves obtained by the laboratories 1, 6, and 7, using odd or wrong calibrations are detected as statistically different from each other and from the ones obtained by laboratories 25. This fact supports the use of functional ANOVA as an alternative in consistency studies for laboratories, as long as the data outputs were curves.

The methodology established by the ASTM E691 is used to compare the corresponding results with those obtained with the proposed new functional approach. The number of outliers identified using a standard ILT depends on the extracted feature (e.g., IDT) used in the analysis. Not all the features summarize properly all the information of a whole thermal curve. Therefore, the use of FDA techniques is justified because the non-consistent laboratories can be identified using the whole curves, before extracting the interest features in each thermal curve.

This procedure can be applied in a wide range of ILT problems where the results can be managed as functional data, as it is the case of curves obtained by thermal analysis techniques: TG, DSC, dynamic mechanical analysis (DMA), thermomechanical analysis (TMA), and dielectric analysis (DEA). In addition, this procedure does not need to fulfill any previous assumption such as data following a Gaussian distribution.

Acknowledgements This research has been partially supported by the Spanish Ministry of Science and Innovation. Grant MTM2011-22392 (ERDF included). The authors thank the two referees for very helpful comments and suggestions.

References

1. Tarrío-Saavedra J, Naya S, Francisco-Fernández M, Artiaga R, López-Beceiro J. Application of functional anova to the study of thermal stability of micro-nano silica epoxy composites. Chemometrics Intell Lab Syst. 2011;105:114–24.
2. 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.
3. 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.
4. Tarrío-Saavedra J, Naya S, López-Beceiro J, Gracia-Fernández C, Artiaga R. Thermooxidative properties of biodiesels and other biological fuels. In: Montero G, Stoycheva M, editors. Biodiesel. Rijeka: INTECH; 2011. p. 47–62.
5. Practice for conducting and interlaboratory study to determine the precision of a test method. Annual Book of ASTM Standards. West Conshohocken, PA: ASTM International E691; 2004.
6. Accuracy (Trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. Geneva: ISO. International Standard ISO 5725-4-1994; 1994.
7. Hund E, Massart DL, Smeyers-Verbeke J. Interlaboratory studies in analytical chemistry. Ann Chim Acta. 2000;423:145–65.
8. Lysiak-Pasttuszak E. Interlaboratory analytical performance studies: a way to estimate measurement uncertainty. Oceanologia. 2004;46(3):427–38.
9. Ramsay JO, Silverman BW. Functional data analysis. New York: Springer-Verlag; 2005.
10. Cuevas A, Febrero M, Fraiman R. On the use of the bootstrap for estimating functions with functional data. Comput Stat Data Anal. 2006;51:1063–74.
11. Fraiman R, Muniz G. Trimmed means for functional data. Test. 2001;10:419–40.
12. Cuevas A, Febrero M, Fraiman R. Robust estimation and classification for functional data via projection-based depth notions. Comput Stat. 2007;22:481–96.
13. Cuevas A, Febrero M, Fraiman R. An anova test for functional data. Comput Stat Data Anal. 2004;47:111–22.

14. Brumback BA, Rice JA. Smoothing spline models for the analysis of nested and crossed samples of curves. J Amer Stat Assoc. 1998;93:961–94.
15. Ramsay JO, Silverman BW. Applied functional data analysis. New York: Springer-Verlag; 2002.
16. Ramsay JO, Hooker GS. Functional data analysis with R and Matlab. New York: Springer; 2009.
17. Fan J, Lin KS. Functional anova models for proportional hazards regression. J Amer Stat Assoc. 1998;93:1007–21.
18. Shen Q, Faraway JJ. An F test for linear models with functional responses. Stat Sin. 2004;14:1239–57.
19. Cuesta-Albertos JA, Febrero-Bande M. A simple multiway anova for functional data. Test. 2010;19:537–57.
20. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2008. http://www.R-project.org.
21. Febrero-Bande M, de la Fuente OM. Statistical computing in functional data analysis: the R package FDA.usc. J Stat Soft. 2012;51:1–28.
22. Febrero-Bande M, Galeano P, González-Manteiga W. Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. Environmetrics. 2008;19:331–45.
23. Brunner E, Dette H, Munk A. Box-type approximations in nonparametric factorial designs. J Am Stat Assoc. 1997;92:1494–502.
24. Miller RG. Simultaneous statistical inference. New York: Springer-Verlag; 1991.
25. Ferraty F, Vieu P. Nonparametric functional data analysis. Berlin: Springer-Verlag; 2006.
26. Tarrío-Saavedra J, López-Beceiro J, Naya N, Artiaga R. Effect of silica content on thermal stability of fumed silica/epoxy composites. Polym Degr Stab. 2008;93:2133–7.