

A Round-Robin Study on Decomposition of Calcium Oxalate Monohydrate by Thermogravimetry

Suzanne, M.^{1,*} · Bourbigot, S.² ·
Carlotti, P.^{1,**} · Coker, E.³ ·
Dumazert, L.⁴ · Duquesnes, S.² ·
Fateh, T.⁵ · Guillaume, E.⁵ ·
Hakkarainen, T.⁶ · Lazaro-Urrutia, D.⁷ ·
Leventon, I.⁸ · Rein, G.⁹ · Richter, F.⁹ ·
Rinta-Paavola, A.¹⁰ · Rogaume, T.¹¹ ·
Schartel, B.¹² · Sjöström, J.¹³ ·
Stoliarov, S.¹⁴ · Thiry, A.¹ ·
Witkowski, A.¹⁵

Manuscript submitted for review

1 Author contacts

Suzanne Mathieu: mathieu.suzanne@interieur.gouv.fr
Bourbigot Serge: serge.bourbigot@centralelille.fr
Carlotti Pierre: pierre.carlotti@arteliagroup.com
Coker Eric: encocker@sandia.gov
Dumazert Loïc: loic.dumazert@mines-ales.fr
Duquesnes Sophie: sophie.duquesne@univ-lille.fr
Fateh Talal: talal.fateh@efectis.com
Guillaume Eric: eric.guillaume@efectis.com
Hakkarainen Tuula: tuula.hakkarainen@vtt.fi
Lazaro-Urrutia David: david.lazaro@unican.es
Leventon Isaac: isaac.leventon@nist.gov
Rein Guillermo: g.rein@imperial.ac.uk

¹ Laboratoire Central de la Préfecture de Police, Paris, France

² Ecole Centrale de Lille, Lille, France

³ Sandia National Laboratories, Albuquerque, NM, USA

⁴ IMT Mines d'Alès, Alès, France

⁵ Efectis, France

⁶ VTT, Espoo, Finland

⁷ GIDAI. Universidad de Cantabria, Spain

⁸ NIST, Gaithersburg, MD, USA

⁹ Imperial College, London, UK

¹⁰ Aalto University, Espoo, Finland

¹¹ Institut Pprime (CNRS), université de Poitiers, Poitiers, France

¹² Bundesanstalt für Materialforschung und Prüfung, Berlin, Germany

¹³ RISE, Göteborg, Sweden

¹⁴ University of Maryland, College Park, MD, USA

¹⁵ University of Central Lancashire, Preston, UK

*Corresponding author: mathieu.suzanne@interieur.gouv.fr

** now at Artelia, Choisy-le-Roi, France

Richter Franz: franz.richter11@imperial.ac.uk
Rinta-Paavola Aleks: aleksi.rinta-paavola@aalto.fi
Rogaume Thomas: thomas.rogaume@univ-poitiers.fr
Schartel Bernhard: bernhard.schartel@bam.de
Sjöström Johan: johan.sjostrom@ri.se
Stoliarov Stanislav: stolia@umd.edu
Thiry Aurélien: aurelien.thiry@interieur.gouv.fr
Witkowski Artur: awitkowski@uclan.ac.uk

A Round-Robin Study on Decomposition of Calcium Oxalate Monohydrate by Thermogravimetry

List of authors

Manuscript submitted for review

Abstract A round-robin study for thermogravimetric analysis (TGA) of carbon oxalate monohydrate was organised with thirteen participants. The main objectives were to provide best practice in the use of TGA and guidelines on uncertainties in the context of pyrolysis and fire modelling. The material studied, calcium oxalate monohydrate, is often used to evaluate the proper functioning of a given piece of equipment because its decomposition steps are well documented and known to be repeatable and reproducible. The experimental procedure was described in as much detail as possible, bearing in mind that it cannot be fully prescribed due to the different technologies of thermogravimetric analysers. All thermograms were evaluated for mass losses and onset, peak and end-set temperatures for the three decomposition steps using two outlier detection methods, the Grubbs method and an alternative two-step method inspired from Z-score technique, allowing an assessment of measurement deviation calculated throughout the duration of the experiment. This study allows a benchmarking of data scatter due to instrument and operator. It confirms that verification with calcium oxalate is a good screening method that can be used by laboratories. Within a laboratory, it can also help to identify problems with instrument accuracy that would otherwise be difficult to detect. This, together with the good practice outlined in the paper, should help the community to increase the validity of tests on other materials of interest to fire science.

Keywords Thermal decomposition · benchmark · Pyrolysis · TGA analysis

1 Introduction

Thermogravimetric analysis (TGA) is an analytical technique for measuring changes in the mass of a material that occur in response to programmed temperature changes [1]. It is used to understand sample composition, predict lifetimes, or study reactions of material with gases. TGA is also widely used in fire research or fire engineering. It allows (a) the study of thermal behaviour of a material [2], potentially loaded with flame retardant or other fillers, or (b) the determination of kinetic parameters

List of institute

Table 1 List of participants to the present round-robin.

Participant 1	City, Country
Participant 2	City, Country
Participant 3	City, Country
Participant 4	City, Country
Participant 5	City, Country
Participant 6	City, Country
Participant 7	City, Country
Participant 8	City, Country
Participant 9	City, Country
Participant 10	City, Country
Participant 11	City, Country
Participant 12	City, Country
Participant 13	City, Country
Participant 14	City, Country

[3] and the extrapolation of a given material behaviour at other heating rates or at larger scales [4]. From 1986 to 2020, several round-robin studies were performed with TGA on cellulose [5,6], wood [6], synthetic polymers [7,8] or carbon fibre-reinforced thermoplastic composites [9]. The material choice was obviously related to the research domain of participants and round-robin organisers. These studies show that a good reproducibility of TGA measurements can be achieved but it strongly depends on the nature of the material. This assessment is valid for a wide range of equipment. However, these comparisons are made on reduced kinetic parameters after data processing and not on raw data. As a consequence, the assessment is less on the TGA experimental process itself than on the whole chain (measurement followed by specific post-processing).

A round robin test on thermogravimetric analysis was therefore organised to validate experimental methods and improve result quality. It gathers enough participants active in fire research to assess the reproducibility of measurements as well as to trigger discussions on best practice. Eighteen participants were initially interested in the project; fourteen sets of measurement were received provided by thirteen entities. One of the participants performed measurements with two types of equipment. Two laboratories were working together and provided only one set of measurements. The list of participants is available in Table 1.

The paper falls into four main parts after this introduction. Section 2 presents the various TGA equipments used, test material used in this round-robin, and the common protocol used by all participants together with the specificities of each participant. Then the results are presented in section 3: temperature programme conformity, mass loss curves, decomposition temperatures. Outlier detection using Grubbs method and then an alternative method is presented in section 4, before some recommendations, which falls into two parts: good practice for TGA and methods for outliers detection in TGA. Section 5 is the conclusion.

2 Experimental protocol

2.1 Equipment

There are several manufacturers of TGA apparatus, but all are based on the same principle: the sample is put in a crucible, which weight is measured continuously along the experiment. The complete system is put in an enclosure heated by an electric resistance, called oven, and swept by a gas flow. The aim is that the actual sample temperature T_s follows precisely a temperature programme as function of time, $T_p(t)$. This is done by using thermocouples closely connected to the sample, and using an active control of power delivered to the heating resistance as function of $T_p(t)$ and measured temperature at the thermocouple T_c . Since there is no actual way of knowing the actual sample temperature, the calibration procedure on reference material is of crucial importance. Generally it uses material with a phase change (either melting point or Curie point) in the temperature range of interest to get insight on actual sample temperature. Similarly, the systems used for weight measurements vary. They can be based on a horizontal arm, top-deck scale or a suspension load cell. Each TGA apparatus follows the above principles, but vary in the way they are designed and built. In the present work, a total of five different brands and ten different TGA apparatus were used. Participants were encouraged to provide as much information as possible. Table 2 summarises the technical specifications of these instruments as provided by the manufacturers.

For the instruments of most participants, mass calibration is performed using an automatic program realised internally. A participant verifies the auto-calibration by performing a test on 15 mg to 20 mg of CaCO_3 in air atmosphere at 20 K min^{-1} from 773 K to 1173 K. Four participants verify their weighting scale with external known masses. As already stated, temperature calibration is critical to ensure a good conformity to temperature programme. Table 3 summarises information regarding the temperature calibration of instruments; the frequency of calibration was provided by most participants although not initially requested in the round-robin procedure.

2.2 Material sample

The material used for the sample was determined to have several qualities: (1) being well studied and with stable manufacturing conditions; (2) having several decomposition steps, well separated in temperature; (3) having known values for residual mass at each step; (4) having a large amount of residue after final step (this last property was chosen to add an extra criterion to assess reproducibility). Thus, a widely known material was chosen for this study: calcium oxalate monohydrate in powder form from Sigma-Aldrich (ref. C0350000, CAS No. 5794-28-5). All the samples used in this study came from the same production batch acquired and dispatched to all the participants. Indeed, the following sequence is commonly reported for the decomposition of calcium oxalate monohydrate into calcium oxalate, then into calcium carbonate and finally into calcium oxide:

$$\text{CaC}_2\text{O}_4 \cdot \text{H}_2\text{O} \longrightarrow \text{CaC}_2\text{O}_4 + \text{H}_2\text{O} \uparrow \longrightarrow \text{CaCO}_3 + \text{CO} \uparrow \longrightarrow \text{CaO} + \text{CO}_2 \uparrow$$

For an initial mass of sample m_0 , and m_i residual mass after step i , we define nor-

Table 2 Characteristics of the instruments

	Brand & Model	Load method	Mass sensitivity	Position and number of thermocouple
A	Netzsch STA 449 F3 Jupiter	Top	0.1 μ g	1 at the bottom of each crucible
B	Mettler Toledo SDTA 851e	Horizontal	n/a	n/a
C	Netzsch Libra	Top	0.1 μ g	Furnace and sample holder
D	Netzsch STA 449 F3 Jupiter	Top	0.1 μ g	1 at the bottom of each crucible
E	Perkin-Elmer Pyris 1	Suspension	0.1 μ g	1 at the bottom of the furnace
F	Setaram Setsys Evolution	Suspension	10 μ g	1 at the bottom of each crucible
G	Mettler Toledo TGA/DSC 1	Horizontal	5 μ g	3 at the bottom of each crucible
H	Netzsch TG 209 F1 Iris	Top	0.1 μ g	1 under the sample
I	Mettler Toledo TGA/DSC 1	Horizontal	5 μ g	3 at the bottom of each crucible
J	Netzsch STA 449 F3 Jupiter	Top	0.1 μ g	1 at the bottom of each crucible
K	Netzsch STA 449 F3 Jupiter	Top	0.1 μ g	1 at the bottom of each crucible
L	TA Instruments TGA Q500	Suspension	0.1 μ g	2 above the sample
M	Mettler Toledo TGA/DSC 2 LF/HT	Horizontal	0.1 μ g	3 at the bottom of each crucible
N	Netzsch STA 449 F1 Jupiter	Top	0.025 μ g	1 at the bottom of each crucible

Table 3 Temperature calibration of instruments; elements are denoted by their symbol in periodic classification and alloys by their commercial names; n/a: non available; (*) number of calibration points depends on crucible; last calibration for participant D was made 7 years before the tests.

	Method	Calib. points	Material	Range (K)	Frequency
A	Melting point	5	In, Sn, Bi, Zc, Al	[429:933]	n/a
B	Melting point	4	In, Sn, Zc, Al	[429:933]	1 year
C	Melting point	5	n/a	n/a	0.5 year
D	Melting point	n/a	n/a	n/a	n/a - see caption
E	Curie point	4	Fe, Ni, AlumeI Perkalloy	[303:1223]	1 year
F	Melting point	n/a	n/a	[303:1373]	1 year
G	Melting point	3	In, Al, Au	[373:1373]	1 year
H	c-DTA method	n/a	n/a	[423:1233]	n/a
I	Melting point	2	In, Al	[429:933]	1 year
J	Melting point	n/a	n/a	[429:1723]	0.25 year
K	Melting point	4 to 6(*)	n/a	[429:1337]	n/a
L	Curie point	2	n/a	[631:933]	1 month
M	Melting point	2	In, Zc	n/a	1 year
N	Melting point	6 to 8(*)	In, Sn, Bi, Zc, Al, Ag	[342:1235]	0.5 to 1 year

malised mass losses for each step as

$$ML_1 = \frac{m_0 - m_1}{m_0} \quad , \quad ML_2 = \frac{m_1 - m_2}{m_0} \quad , \quad ML_3 = \frac{m_2 - m_3}{m_0} \quad (1)$$

Using molecular masses, one may compute their theoretical values:

$$ML_1^T = 12.3\% \quad , \quad ML_2^T = 19.2\% \quad , \quad ML_3^T = 30.1\%$$

They are consistent with data in the literature [10].

2.3 Experimental protocol

One purpose of the present work is the assessment of relative scatter in data due to the different equipment. The experimental procedure was therefore as detailed as possible in order to allow as many different apparatus while still providing guidelines. When needed, this experimental procedure took into account the technological differences of the aforementioned testing equipment. Furthermore, each participant also relied on internal quality procedures adapted to its own context.

The sweeping gas had to be inert with regards to calcium oxalate and its derivatives. All participants used nitrogen except one, who employed argon instead (see Table 4). The actual value of flow rate of inert gas is dependent on the equipment and was therefore not prescribed in the experimental protocol. This flow rate can sometimes affect the measurement [?,11]. The effect of the flow rate of the inert gas on the experimental mass loss curves for the thermal dehydration of calcium oxalate monohydrate for a horizontally supported thermogravimetric analyser has been studied in [12]. It is indicated that, for a reversible reaction, an increase in the flow rate would promote the removal of the gaseous products, and therefore, would accelerate the reaction shifting its temperature range to lower temperatures. It is recommended to maintain a sufficiently fast flow rate of a purge gas in order to maximise the efficiency of the removal of gaseous products. As a consequence participants had to ensure that this parameter did not influence their measurements as specified in [12]. At the beginning of each test, the whole experimental set-up had to be purged with the chosen inert gas for 1200 s to 1800 s before starting the actual heating program [12], to ensure that no oxygen was present in the oven when experiments started without purge.

Concerning initial mass of sample, standards recommend sample mass values over 10 mg with a maximum specified value of 100 mg [13] and 30 mg [14]. However, this recommendation is highly discussed in the literature. For a heating rate of 10 K min⁻¹, [2] mentions that a smaller sample, of size of 5 mg to 10 mg, is needed to ensure that it is in thermal equilibrium with the apparatus. These authors indeed recommend to keep the sample size as small as possible within the limits of sensibility of the apparatus which usually leads to around 5 mg. A rule of thumb indicated in [12] recommends that the sample mass times heating rate should not exceed 100 mg K min⁻¹. This value, which corresponds to 10 mg at 10 K min⁻¹ was obtained from a Biot number calculation [12]. This assumption might however not be entirely verified as suggested by several publications where no convergence is observed for PP even for sample masses as small as 0.14 mg [15], 0.83 mg (at 15 K min⁻¹) [16] or 0.11 mg (at 40 K min⁻¹) [5]. From the above considerations, sample initial mass of 5 mg and heating rate of 10 K min⁻¹ were demanded, with temperatures ranging

Table 4 Inert gas and flow rates and nature and characteristics of crucibles; note that for some participants, two flow-rates are reported: the main flow-rate plus an additional inert gas flow-rate for weighing scale protection.

	Gas	Flow rate (mL min ⁻¹)	Crucible material	Diameter (mm)	Height (mm)	Volume (μ L)	Mass (mg)
A	Nitrogen	50	Alumina	6.8	3.9	n/a	n/a
B	Nitrogen	50	Alumina	6	4.6	70	180.2
C	Nitrogen	40	Alumina	6.8	4	85	140
D	Nitrogen	30	Alumina	n/a	n/a	900	1282
E	Nitrogen	20	Ceramic	7	2.5	20	140
F	Nitrogen	40	Alumina	5	25	300	450
G	Nitrogen	20+80	Alumina	6	4.5	70	181
H	Nitrogen	20+10	Alumina	6.8	n/a	85	n/a
I	Nitrogen	80	Alumina	5	4.5	70	172
J	Argon	100	Alumina	6	4	100	152.7
K	Nitrogen	40	Alumina	6.8	4	85	147.5
L	Nitrogen	60+40	Alumina	10	4	250	632.8
M	Nitrogen	40	Alumina	n/a	n/a	n/a	n/a
N	Nitrogen	20+30	Alumina	5.9	3.7	40	138

from 300 K to 1180 K. The protocol also asked for 300 s isotherms at the beginning and end of each test. All participants complied with the 10 K min⁻¹ heating rate. They also complied with the 5 mg initial sample mass (5.12 mg $\pm$ 0.19 mg) except for participant C which started the experiment with a mass of 20.45 mg.

Crucibles were asked to be made of alumina, with no indication for their sizes. The sizes were not imposed because they are strongly dependent on the equipment characteristics. Crucible size may nevertheless influence the measurement. Indeed, it is better for a given sample mass to compare measurements performed on the same thickness of material to have similar diffusion rates of volatile decomposition products [12]. Crucibles were indeed composed of alumina for all participants except one participant who also used a ceramic but without specifying its nature (see Table 4). It is interesting to note that participant D used a 1282 mg crucible, which is almost 10 times heavier than the crucible used by some of the other participants.

No conditioning was requested for the sample in the experimental protocol, because it was thought it is dependant on each laboratory quality procedures, conditioning equipment and experience. Seven of the participants reported nevertheless to have carried out conditioning of the sample before performing measurements (see Table 5). Participants were invited to repeat each test three times but it was not mandatory. Three repeats were performed for nine sets of data.

3 Experimental results

3.1 Temperature programme conformity

First, the temperature programme conformity is analysed. All participants reported sample temperature T_s as a standard output of their apparatus. As presented in Table 2, the number and location of thermocouples depend on the apparatus model, and the critical assessment of whether T_s is actually representative of the sample temperature is not possible. However, it is significant to use these data, as they are designed

Table 5 Conditioning of calcium oxalate monohydrate before TGA measurements

	Duration	Temperature (K)	Relative humidity (%)
A	Minimum of 48 h in desiccator with drierite	n/a	n/a
B	n/a	n/a	n/a
C	12 h	353	n/a
D	n/a	n/a	n/a
E	n/a	n/a	n/a
F	n/a	n/a	n/a
G	1800 s in TGA oven	300	n/a
H	Minimum of 48 h	269.15	50
I	Minimum of 48 h in desiccator with drierite	n/a	n/a
J	n/a	n/a	n/a
K	Minimum of 48 h	n/a	n/a
L	n/a	n/a	n/a
M	n/a	n/a	n/a
N	Minimum of 48 h	291	15

to follow exactly the temperature programme as function of time $T_p(t) = \beta_p t$ during the heating phase. In order to analyse programme conformity, the estimated sample heating rate was therefore computed

$$\beta_s = \frac{dT_s}{dt} \quad (2)$$

and compared with $\beta_p = 10$ K/min. For simplicity of reading, β_s is plotted as function of T_s in Fig. 1. Note that taking derivative always generates a risk of increasing the noise because of over-sampling or under-resolved data. Therefore the sampling rate of dataset was first decreased to insure that temperature differences were actually significant. After this stage, only one dataset (dataset D) was still under-resolved and therefore noisy. Bernstein polynomials smoothing was applied to dataset D (see [17,18] for general considerations on the method and [19] for an example of its application to fire science). This smoothing method is particularly effective for high frequency noise, such as that observed for this participant. It is therefore considered that the smoothing does not affect the data set and its use in the rest of this study.

The curves in Fig. 1 are typical of active control with devices as Proportional-Integral-Derivative (PID) controllers with, in most cases, a transient over- or under-shoot and in one case an oscillatory behaviour. Results can be sorted in three categories: (1) participants with an over or undershoot which is smaller than 3 K/min above 350 K and smaller than 1 K/min above 400 K, hence following very well the temperature programme in the range of temperatures of interest; (2) participants with a stronger overshoot, which disappear around 500 K; (3) one participant with an oscillatory behaviour slowly damped but still present at 800 K. These facts have to be interpreted with care: since the mass loss data are plotted as function of temperature, the first order effect of non conformity in temperature programme is excluded, and the remaining effect is second order: transient heat storage in the samples and the crucibles. Therefore, it is not possible from these data to decide whether a measurement should be excluded or not. A thorough analysis of mass loss curves is needed to go further.

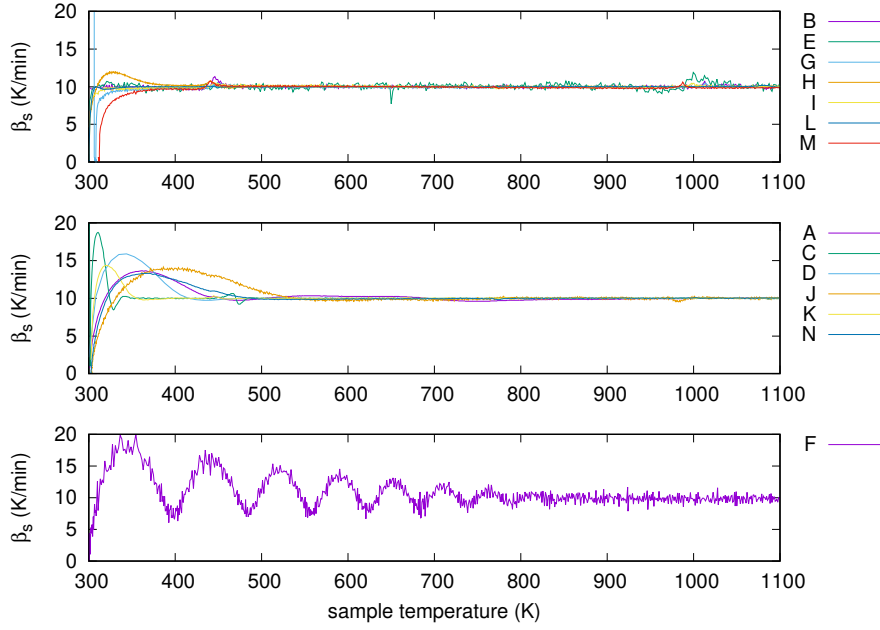


Fig. 1 Temperature programme conformity: β_s plotted as function of sample temperature T_s ; top panel: participants with conformity achieved below 400 K (see text); middle panel: participants with non oscillatory behaviour, but conformity reached slowly (note that a smoothing had to be practised for participant D, see text); bottom panel: participant with oscillatory behaviour.

3.2 Mass loss data

As stated previously, the majority of participants performed several repeats. Except for participant F, it appears visually that an excellent repeatability is reached (not shown here as curves collapse perfectly). The only exception is participant F who had a real difficulty of repeatability for step 1 (see Table 6). For the rest of the paper, only repeat 2 was kept for participant F.

Normalised masses (i.e. $m(T)/m_0$ for each test) are plotted in Fig. 2 as functions of sample temperature for the fourteen sets of data. A general tendency is clearly observable but discrepancies are visible for mass losses as well as for onset or endset temperatures. These discrepancies tend to accumulate one decomposition step after the other, to be finally much more visible for the third decomposition step.

The participants determined from their measurements the mass lost during the three decomposition steps. For data sets with repeats, an average value was first calculated for each participant and this value later used in the determination of the overall average. Normalised mass losses are given in Table 6.

For each decomposition step, the average normalised mass loss is lower than the theoretical normalised mass loss. This means that at each step the normalised residual mass is larger than its theoretical value. This might be caused by incomplete reaction, however this is not likely from the data since the mass residuals appears to be well stabilised. The most probable explanation would be the presence of additive

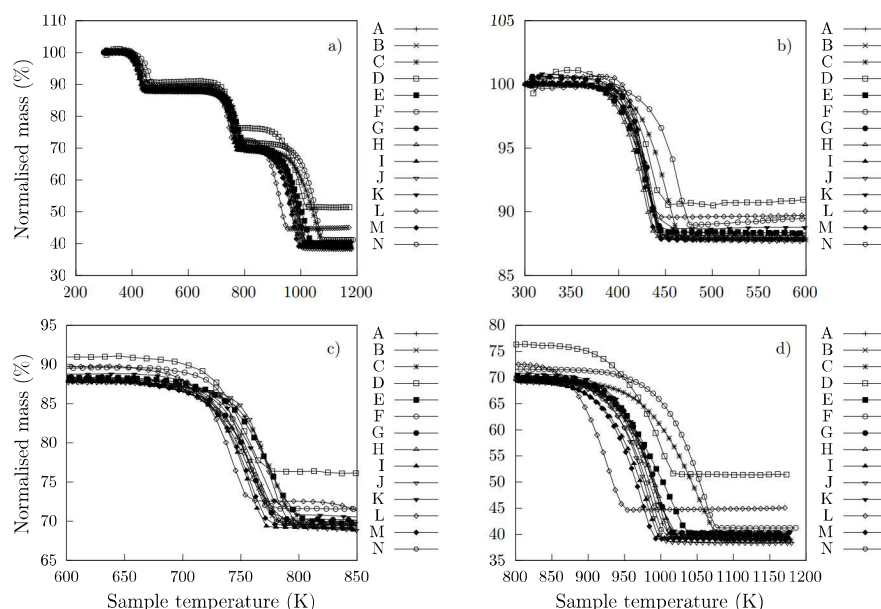


Fig. 2 Normalised mass as a function of sample temperature – Comparison between participants including all dataset

in the samples. However, since all samples were taken from the same batch of production, this does not raise concern as far as this round robin study is concerned. Furthermore, as expected for increasing decomposition steps, it is observed an increase in the standard deviation as well as in the difference between the maximal and minimal measurements. This can be seen as a build-up of measurement errors from previous decomposition steps.

3.3 Decomposition temperatures

Participants had to determine from their measurements the onset, peak and end-set temperatures for each decomposition step, using their own methods. For data sets with repeats, an average value was first calculated and later used in the determination of an overall average. Unlike for the mass lost during a decomposition step (see section 3.2), no theoretical value exists as a reference for the decomposition temperature, which depends on the experimental procedure (*e.g.* initial sample mass or heating rate).

Data are also plotted in Fig. 3 where temperatures reported for the 1st step are in black with no fill, cross-hatched for the 2nd step and hatched for the 3rd step. The bottom line of each rectangle corresponds to the onset temperature, the centre line to the peak temperature and the top line to the end-set temperature. These temperatures are also compared to the averaged onset, peak and end-set temperatures plotted with their standard deviations on the right hand side of the figure.

Table 6 Normalised mass loss ML_1 , ML_2 and ML_3 (in %) as defined by eq. (1); note that the largest uncertainty in mass measurement for F is $10\mu\text{g}$ representing 0.2 % of initial mass; uncertainties for G and I is 0.2 % of initial mass, and less than 0.002 % of initial mass for A, C, D, H, J, K, L, M and N (see Table 2). R-1 to R-3 correspond to the test replicates, if applicable.

	ML_1	ML_2	ML_3
A	12.8	18.0	30.0
B	11.8	18.1	29.9
C	12.2	18.7	30.3
D	9.3	12.6	20.8
E-R1	11.7	18.6	29.9
E-R2	11.7	18.6	29.8
E-R3	11.8	18.9	30.0
F-R1	11.1	18.3	29.9
F-R2	10.7	18.0	30.2
F-R3	10.1	18.0	31.3
G-R1	12.2	18.2	29.9
G-R2	12.3	18.1	29.9
G-R3	12.2	18.2	29.7
H	12.1	18.4	30.3
I-R1	12.0	19.0	30.0
I-R2	12.0	19.0	30.0
I-R3	12.0	19.0	30.0
J-R1	12.3	18.7	30.6
J-R2	12.4	18.5	30.2
J-R3	12.3	18.5	30.3
K-R1	11.2	18.3	29.9
K-R2	11.2	18.3	29.9
K-R3	11.2	17.9	30.1
L	11.0	17.2	27.9
M	12.0	18.0	30.0
N-R1	12.1	18.5	29.7
N-R2	12.1	18.4	30.0
N-R3	12.1	18.5	30.1
Theory [10]	12.3	19.2	30.1
Median	12.0	18.2	30.0
Average	11.6	17.9	29.3
Standard deviation	0.8	1.5	2.4

To analyse these results, a comparison is made between measurements from participants A and G which are very close (see Fig. 2). The lengths of the three rectangles from participant A in Fig. 3 are however much bigger than those of participant G. For the first, second and third decomposition steps, the rectangle lengths for participant A are respectively 52 K, 82 K and 110 K, whereas for participant G these rectangle lengths are respectively 32 K, 47 K and 60 K. Both participants did not use the same standard or protocol to determine the onset and end-set temperatures which led to such differences. On the contrary, the reported peak temperatures of these participants were much closer: an absolute difference of 2 K for the first decomposition step between peak temperatures of participants A and G, 4 K for the second decomposition step and 16 K for the third decomposition step. The determination of peak temperature appears to be less dependent on the chosen standard or method. Finally, as observed previously in section 3.2, the standard deviations reported in Table 6 in-

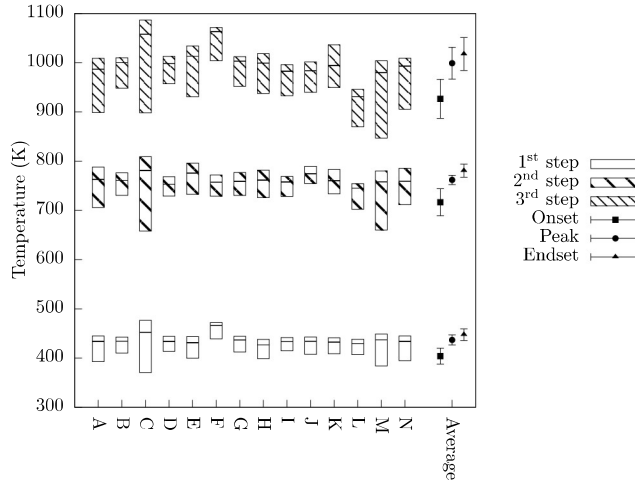


Fig. 3 Onset, peak and endset temperatures for the three decomposition steps compared with average values

crease with increasing decomposition steps due to an accumulation of measurement errors.

4 Discussion

4.1 Detection of outliers: Grubbs test method

4.1.1 Systematic tests for outlier detection

From previous section, it is clear that a group of results are well concentrated while other are further away and can be seen as outliers. The aim of the present section is to apply a generic method for outlier detection and see whether general conclusions can be drawn from the analysis. The method chosen for outlier detection is Grubbs test [20, 21] which is generally recommended for small data sets (*i.e.* less than 21 test results). The principle of Grubbs test for a series of data is the following.

Starting with n data, $x_1, x_2, \dots, x_n$ sorted in increasing order. First, compute V_n , the variance of the full set of data, and V_{n-1} the variance of data excluding the largest one. Assuming data follow Gaussian pdf, statistical theory shows that V_{n-1}/V_n has to be close to one with a known probability. Hence, for outlier detection, it is decided to exclude or not the largest data depending on the value of V_{n-1}/V_n compared with a tabulated value depending on the number of data and the significance threshold retained. The procedure is then carried out similarly for the smallest of the data, and so on iteratively until no outlier is found. One of the known dangers of Grubbs test is over-exclusion, since variance is recalculated at each step after having excluded a new outlier; thus, an overly stringent confidence level should not be used.

Table 7 Statistics of normalised mass loss ML_1 , ML_2 and ML_3 (in %) as defined by eq. (1) excluding outliers D, F, K and L.

	ML_1	ML_2	ML_3
Theory [10]	12.3	19.2	30.1
Median	12.1	18.45	30.0
Average	12.0	18.4	30.1
Standard deviation	0.2	0.4	0.2

We nevertheless decided to use Grubbs test because of its generality and broad application. The confidence level retained in the present work is 2.5 %. Since the process is iterative and monovariate, we decided to work first on mass loss and then on temperature on the remaining set.

4.1.2 Mass loss data

From the mass loss data with 2.5 % confidence level, results from participants D, F, K and L are then detected as outliers, leaving us with 10 datasets: A, B, C, E, G, H, I, J, M, and N.

As mentioned previously, crucibles of participants D and L were respectively 8 times and 4 times heavier than those of other participants which may explain these observed differences in mass measurements. Indeed, as the sample mass was set at 5 mg in the experimental protocol, it is suspected that these instruments did not provide good data because the prescribed sample mass is outside their operating conditions or because of an issue with the baseline correction/conditioning step, which later propagated into these difference in apparent mass loss in each step. For participants F and K, they are considered as outliers due to a poor representation of the first decomposition step.

The average value without participant data detected as mass outliers is compared to the theory in table 7. These data can be directly compared to those listed in table 6 which contain measurements from all the participants. The exclusion of mass outliers significantly reduces the standard deviation which is lower than 0.35% for the third decomposition step compared to 2.4% with mass outliers. The difference between the maximal and minimal values is divided by up to 16 for the third decomposition step. And more importantly, the relative differences between the average and theoretical values are reduced from 5.7 %, 6.8 % and 2.7 % for respectively the 1st, 2nd and 3rd decomposition steps to 2.4 %, 4.2 % and 0 %. Normalised masses as functions of sample temperature are plotted in Fig. 4 without data from participants detected as mass outliers. The dispersion of mass measurements is smaller along the vertical axis than the one observed in Fig. 2. However, a wide dispersion is still visible along the horizontal axis, especially for participant C. which means that some data sets are more precise for mass measurements than for temperature measurements, and show the need to detect also temperature outliers.

4.1.3 Temperature data

Grubbs tests were applied to all datasets, retained the datasets which were identified as mass outliers. Each decomposition step was considered separately. All the participants provided measurements which were compliant with at least eight of the nine

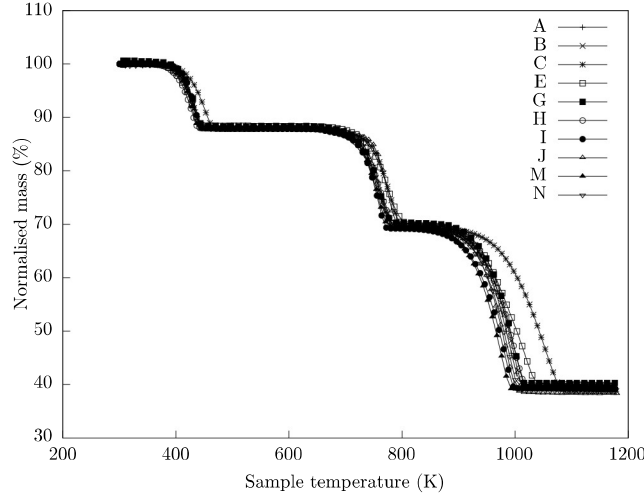


Fig. 4 Comparison between participants of the normalised mass as a function of sample temperature (without measurements having mass outliers); letters A to N refer to dataset presented on table 2.

Table 8 Onset, peak and end-set temperatures (expressed in K) after outliers exclusion (see text) for each decomposition step; T_i^X is for step i with $X = O$ for Onset, $X = P$ for Peak, and $X = E$ for end-set.

	T_1^O	T_1^P	T_1^E	T_2^O	T_2^P	T_2^E	T_3^O	T_3^P	T_3^E
Min	393.0	431.3	441.1	706.0	753.2	768.3	899.0	982.7	996.0
Average	406.1	433.8	443.4	728.6	762.6	781.4	935.1	995.2	1013.6
Max	414.7	436.7	445.0	754.4	775.7	796.0	957.4	1013.3	1036.3
Std. dev.	8.2	1.5	1.4	13.8	7.5	9.5	20.6	10.0	13.4

investigated temperatures (*i.e.* onset, peak and end-set temperatures of the three decomposition steps). Datasets F, C, H, L and M were detected consecutively as outliers in the iterative process. Figure 5 presents temperature data without participant measurements containing temperature outliers, to be compared with Fig. 3. Table 8 gathers statistics for each decomposition step. Standard deviations are drastically reduced: they are divided by two for onset temperatures, between 1.3 and 6.8 for peak temperatures and between 1.4 and 8.4 for end-set temperatures.

Mass measurements as functions of time are plotted in Fig. 6 without participant data detected as temperature outliers. Compared to Fig. 2, measurements in Fig. 6 are less scattered along the horizontal temperature axis. It is interesting to observe that data from participant D which were detected as outliers for mass measurement are not for temperature measurement. This is why this thermogram seems to be detached from other measurements in Fig. 6.

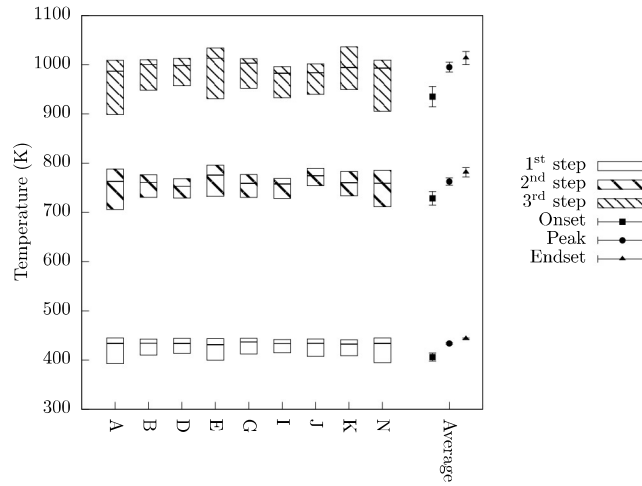


Fig. 5 Onset, peak and end-set temperatures without temperature outliers for the three decomposition steps compared with average values

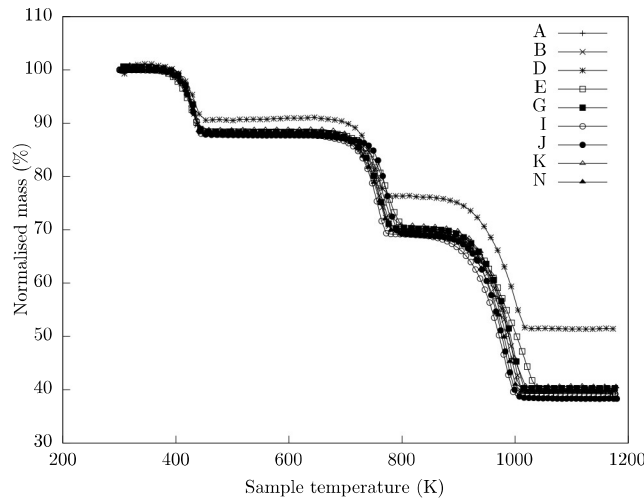


Fig. 6 Comparison between participants of the normalised mass as a function of sample temperature (without measurements having temperature outliers)

4.1.4 Combined mass and temperature outliers

As a whole, D, F, K and L are identified as mass outliers, and F, C, H, L and M as temperature outliers. F and L are double outliers. We are left with 7 non-outliers

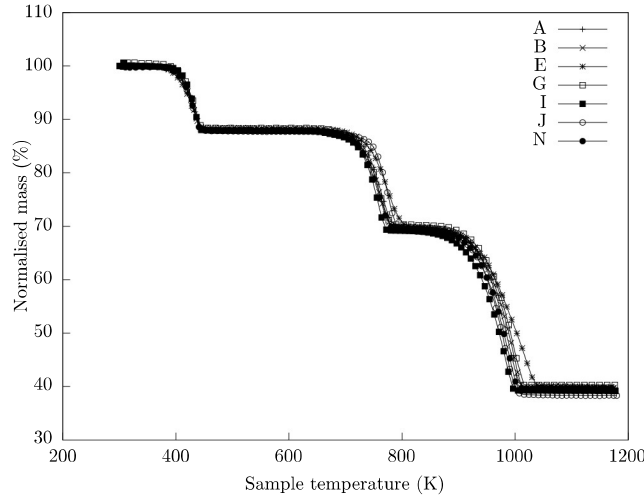


Fig. 7 Normalised mass as a function of sample temperature without participants data with identified mass and temperature outliers

Table 9 Min, Average, Max, and Standard deviation for mass loss (expressed in %), onset, peak and end-set temperatures (expressed in K) for the three decomposition steps without outliers, retaining only A, B, E, G, I, J and N.

	ML_1	ML_2	ML_3	T_1^O	T_1^P	T_1^E	T_2^O	T_2^P	T_2^E	T_3^O	T_3^P	T_3^E
Min	11.7	18.0	29.9	393	431	442	706	758	769	899	983	996
Ave.	12.1	18.4	30.1	405	434	444	728	764	783	930	995	1010
Max	12.8	19.0	30.5	414	437	445	754	776	796	952	1013	1034
St.d.	0.34	0.34	0.24	8.1	1.5	1.2	15	7	9	19	11	11

datasets: A, B, E, G, I, J and N. Figure 7 gives mass loss data from these remaining datasets and Table 9 the associated statistics.

4.2 Alternative method for outlier detection

From the above outlier detection using Grubbs method, one can argue that only 7 participants (A, B, E, G, I, J, and N) out of 14 provided data correct for both mass loss and temperature. However, the method chosen in the present study to select the measurements has two potential limitations. First, Grubbs test works from brute force with a chosen confidence level (2.5 %) which might be too harsh, especially due to the iterative nature of the test which recalculated statistics at each step. Second, the analysis was based on mass losses and onset, peak and end-set temperatures for the three decomposition steps provided by the participant using their own estimation methods. Thus, the analysis detects non exactly measurement outliers, but measurement and processing method outliers. An incorrect estimation of any of the requested mass losses or temperatures may sanction the entire dataset even if the measurement itself was correct. It would be a much better strategy to exclude

outliers from full mass-temperature relationship. However, there is no systematic method for doing this.

We decided to use a two step method inspired from the Z-score method defined in standard ISO13528 [22]. At each step, we compute average and standard deviation as function of temperature using n selected datasets:

$$\langle m \rangle (T) = \frac{1}{n} \sum_i m_i(T) \quad (3)$$

$$\sigma(T) = \sqrt{\frac{1}{n} \sum_i (m_i(T) - \langle m \rangle (T))^2} \quad (4)$$

and finally define

$$Z_i(T) = \frac{m_i(T) - \langle m \rangle (T)}{\sigma(T)}. \quad (5)$$

We then plot the graph $Z_i(T)$ and exclude dataset which have a significant range of temperature with $|Z_i| > 2$. This value corresponds to a 5% percentile criterion for a Gaussian variable, which is common in such studies. For the first step, we keep all the dataset, leading to Fig. 8(a). This leads to the exclusion of datasets D, F, and K. For the second step, average and standard deviation are re-computed after the exclusion, and functions $Z_i(T)$ are computed again. This leads to Fig. 8(b). Now it can be seen that C is an outlier for the third step, while L and N are on the edge, and all the other dataset are well grouped in the middle of the range.

4.3 Recommendations

From the above analysis, it can be seen that, from a Grubbs test point of view, half of the participants would be considered as outliers. On the other hand, a more detailed analysis based on methodology inspired from Z-score would lead to exclude 4 participants out of 14. Recommendations should be grouped into two categories: recommendations of good practice when performing TGA, and recommendations for outliers detection.

4.3.1 Good practice for performing TGA analysis

From the Grubbs analysis, it was complicated to identify reasons for dataset to be excluded. On the other hand, from the second analysis, a few points may be raised.

First, going back to Fig. 1, it appears that F had a real issue of oscillating sample temperature, while C, D and K were in the middle panel showing a strong overshoot. It is unclear whether a more frequent calibration could have brought better results. Calibration periodicity for participant K was unknown, F was one year and D was several years (see Table 3).

Concerning the initial mass of the sample, only C was different from the other participants, and it may also explain the difference of result. This advocate to follow [12] recommendation to keep samples as small as reasonably possible. The crucible size may also be analysed. The fact that D and F used a relatively heavy crucible could be seen as a clue, see Table 4. However, C and K crucibles were well in the range of other tests, while L was the second heaviest, at the edge of in- and outliers. Nevertheless,

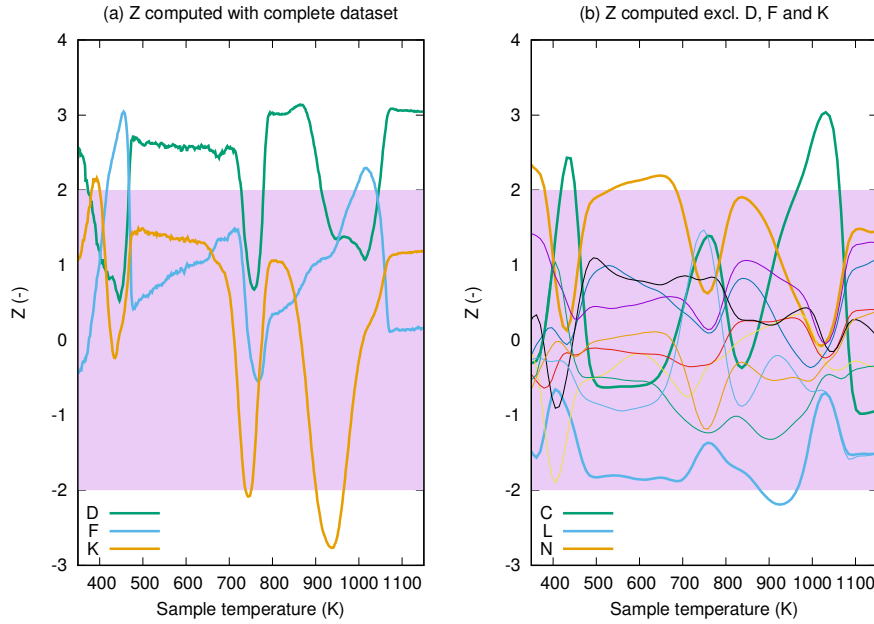


Fig. 8 Generalised Z-score (see text); (a) all dataset are included in the computation of average and standard deviation but only D, F and K are shown; (b) datasets D, F, and K are excluded in the computation of average and standard deviation; all datasets excluding D, F, and K are shown, and lines are smoothed for improved visualisation – only linestyles for C, L and N are specified.

it seems that the effect of crucible mass is poorly understood, and the study of the influence of inert mass in TGA may be of interest in order to investigate scattering among test results.

Finally, performing three repeats serves as a good self-verification practice. Most participants achieved very good repeatability indicating that it is possible to get consistent results from this material. However, even with an excellent repeatability, measurements can deviate from theoretical values.

4.3.2 Outlier detection

Comparing Grubbs method and the two-step method, it seems that the former may sometimes lead to a too high selectivity. In the context of round robin procedure, the main advice here seems limited to a careful inspection of data before using a systematic iterative procedure.

Now, the issue may be formulated in another manner. Suppose a given laboratory performs a TGA in conditions similar to the ones presented here. How will it be possible to analyse these new data in the light of the present ones? For this the answer is double, as described now. First, ML_1 , ML_2 , ML_3 , T_1^O , T_1^P , T_1^E , T_2^O , T_2^P , T_2^E , T_3^O , T_3^P , and T_3^E issued from the new analysis should be compared with data given in Table 9. If all fall in the range given by average plus or minus two times standard deviation, the new data are deemed to be satisfactory according to

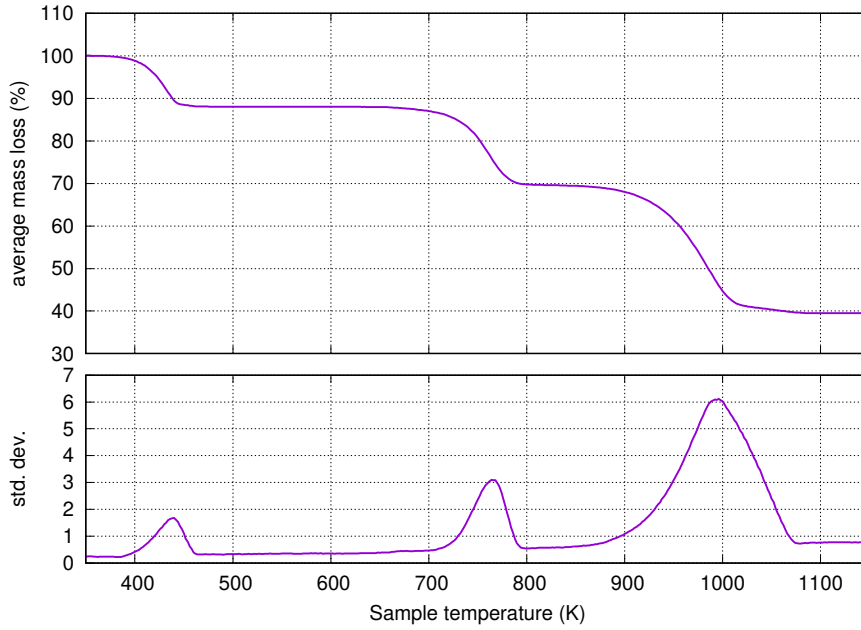


Fig. 9 Normalised mass loss curve (%) as function of sample temperature (K); top pannel: average of all datasets except D, F and K; bottom pannel: standard deviation of all datasets except D, F and K.

common practice (see also [13]). Second, mass loss curve data should be compared with average and standard deviation excluding D, F and K as described in section 4.2. The computation of $Z(T)$ allows to decide whether the new dataset is satisfactory with regards to the alternative test. Average and standard deviation excluding D, F, and K are plotted on Fig. 9 and available on request to the corresponding author.

5 Conclusion

A round-robin test was organised on calcium oxalate monohydrate with thirteen participants with a high level of expertise in the fire domain (leading to fourteen dataset as explained in introduction). The main purpose of this work was the assessment of scatter in data due to equipment. The experimental procedure was consequently as detailed as possible even though it cannot have been entirely prescribed because of the different technologies of thermogravimetry analysers. The investigated material is frequently used to evaluate the proper functioning of a given equipment because its decomposition steps are well-documented and known to be repeatable and reproducible. During the presentation of the results, it was made apparent that temperature programme conformity is a real issue.

The main results of the paper are in section 4. First, using Grubbs method for outliers detection, statistics presented in Table 9 may be used for a stringent test of new TGA runs with calcium oxalate and associated temperature of reaction identification algorithms. Then, in the same context, for the detection of outliers in the

mass loss curves themselves, ideas of section 4.2 apply. Thus, this work confirms that a verification with calcium oxalate is a good screening method, which may be used by laboratories. Within a laboratory it can also assist in finding problems with equipment accuracy otherwise not possible to distinguish. This, together with good practice outlined in section 4, can serve as a validation dataset for the experimentalist, and should help the community to increase the validity of tests with other materials of interest for fire science.

Acknowledgements This work was supported by the International Forum of Fire Research Directors which validated the material selection and experimental procedure. The FORUM (www.fireforum.org) is a group of leaders of fire research organizations from around the world that meet on an annual basis to discuss emerging fire related research topics throughout the world. It has been active since 1991. Its aim is to reduce the burden of fire (including the loss of life and property, and effects of fire on the environment and heritage) through international cooperation on fire research. An important membership requirement is that the applicant be responsible for establishing research priorities and allocating resources for fire safety research in their parent organization. Authors are grateful for their help and recommendations.

Competing interests

G. Rein is the Editor-in-Chief of Fire Technology but was blinded to this peer-review process (which is double-blinded), and the manuscript was managed independently by another editor.

References

1. Sergey Vyazovkin. Thermogravimetric Analysis. In Elton N. Kaufmann, editor, *Characterization of Materials*, page com029.pub2. John Wiley & Sons, Inc., Hoboken, NJ, USA, October 2012.
2. Artur Witkowski, Anna A. Stec, and T. Richard Hull. Thermal Decomposition of Polymeric Materials. In Morgan J. Hurley, Daniel Gottuk, John R. Hall, Kazunori Harada, Erica Kuligowski, Milosh Puchovsky, José Torero, John M. Watts, and Christopher Wieczorek, editors, *SFPE Handbook of Fire Protection Engineering*, pages 167–254. Springer New York, New York, NY, 2016.
3. Sergey Vyazovkin, Alan K. Burnham, José M. Criado, Luis A. Pérez-Maqueda, Crisan Popescu, and Nicolas Sbirrazzuoli. ICTAC Kinetics Committee recommendations for performing kinetic computations on thermal analysis data. *Thermochimica Acta*, 520(1-2):1–19, June 2011.
4. Thomas Rogaume. Thermal decomposition of solid fuels. objectives, challenges and modelling. *Journal of Physics: Conference Series*, 1107:022001, 11 2018.
5. Morten Grønli, Michael Jerry Antal, and Gábor Várhegyi. A Round-Robin Study of Cellulose Pyrolysis Kinetics by Thermogravimetry. *Ind. Eng. Chem. Res.*, 38(6):2238–2244, June 1999.
6. Andrés Anca-Couce, Christos Tsekos, Stefan Retschitzegger, Francesco Zimbardi, Axel Funke, Scott Banks, Tzoulia Kraia, Paula Marques, Robert Scharler, Wiebren de Jong, and Norbert Kienzl. Biomass pyrolysis TGA assessment with an international round robin. *Fuel*, 276:118002, September 2020.
7. Michihiko Momota, Takeo Ozawa, and Hsaaki Kanetsuna. Thermal analysis of polymer samples by a round robin method. *Thermochimica Acta*, 159:125–137, January 1990.
8. T. Ozawa. Industrial standardization of thermal analysis in Japan. *Journal of Thermal Analysis*, 40(3):1379–1398, September 1993.
9. Masahiro Funabashi, Fumi Ninomiya, Akihiro Oishi, Wataru Mizuno, Kimitaka Tahira, Shogo Tanaka, and Masao Kunioka. Round Robin Tests to Determine Fiber Content of Carbon Fiber-Reinforced Thermoplastic Composites by Combustion and Thermogravimetry. *Journal of Polymers*, 2017:1–10, 2017.

10. H.L Anderson, A Kemmler, G.W.H Höhne, K Heldt, and R Strey. Round robin test on the kinetic evaluation of a complex solid state reaction from 13 european laboratories. part 1. kinetic tg-analysis. *Thermochimica Acta*, 332(1):33–53, 1999.
11. David Lázaro, Mariano Lázaro, Alain Alonso, Pedro Lázaro, and Daniel Alvear. Influence of the STA boundary conditions on thermal decomposition of thermoplastic polymers. *J Therm Anal Calorim*, 138(4):2457–2468, November 2019.
12. Sergey Vyazovkin, Konstantinos Chrissafis, Maria Laura Di Lorenzo, Nobuyoshi Koga, Michèle Pijolat, Bertrand Roduit, Nicolas Sbirrazzuoli, and Joan Josep Suñol. ICTAC Kinetics Committee recommendations for collecting experimental thermal analysis data for kinetic computations. *Thermochimica Acta*, 590:1–23, August 2014.
13. ISO11358-1:2014. Plastics Thermogravimetry (TG) of polymers Part 1: General principles, 2014.
14. ASTM E1131-20. *Standard Test Method for Compositional Analysis by Thermogravimetry*. ASTM International, West Conshohocken, PA, 2014.
15. Aurélien Thiry-Muller. *Modélisation de la décomposition thermique des solides*. PhD thesis, Université de Lorraine, 2018.
16. Dawid Stawski. The effect of sample weight in thermogravimetric analysis of low viscosity polypropylene in air atmosphere. *Polymer Testing*, 28(2):223–225, April 2009.
17. J. Dieudonné. *Calcul infinitésimal*. Hermann, 1968.
18. A. Blejec. Bézier curves: simple smoothers of noisy data. In A. Ferligoj and A. Kramberger, editors, *Developments in statistics and methodology*. 1999.
19. A. Suzanne and P. Carlotti. Heat release rate of burning luggage in trains. In M. Tabarra, editor, *International Symposium on Aerodynamics and Fire in Vehicle Tunnels (ISAVFT)*. 2022.
20. Frank E. Grubbs. Sample criteria for testing outlying observations. *Annals of Mathematical Statistics*, 21(1):27–58, 1950.
21. Frank E. Grubbs and Glenn Beck. Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14(4):847–854, 1972.
22. ISO13528. Statistical methods for use in proficiency testing by interlaboratory comparison, 2015.

A Round-Robin Study on Decomposition of Calcium Oxalate Monohydrate by Thermogravimetry

List of authors

Manuscript submitted for review

Abstract A round-robin study for thermogravimetric analysis (TGA) of carbon oxalate monohydrate was organised with thirteen participants. The main objectives were to provide best practice in the use of TGA and guidelines on uncertainties in the context of pyrolysis and fire modelling. The material studied, calcium oxalate monohydrate, is often used to evaluate the proper functioning of a given piece of equipment because its decomposition steps are well documented and known to be repeatable and reproducible. The experimental procedure was described in as much detail as possible, bearing in mind that it cannot be fully prescribed due to the different technologies of thermogravimetric analysers. All thermograms were evaluated for mass losses and onset, peak and end-set temperatures for the three decomposition steps using two outlier detection methods, the Grubbs method and an alternative two-step method inspired from Z-score technique, allowing an assessment of measurement deviation calculated throughout the duration of the experiment. This study allows a benchmarking of data scatter due to instrument and operator. It confirms that verification with calcium oxalate is a good screening method that can be used by laboratories. Within a laboratory, it can also help to identify problems with instrument accuracy that would otherwise be difficult to detect. This, together with the good practice outlined in the paper, should help the community to increase the validity of tests on other materials of interest to fire science.

Keywords Thermal decomposition · benchmark · Pyrolysis · TGA analysis

1 Introduction

Thermogravimetric analysis (TGA) is an analytical technique for measuring changes in the mass of a material that occur in response to programmed temperature changes [1]. It is used to understand sample composition, predict lifetimes, or study reactions

Table 1 List of participants to the present round-robin.

Participant 1	City, Country
Participant 2	City, Country
Participant 3	City, Country
Participant 4	City, Country
Participant 5	City, Country
Participant 6	City, Country
Participant 7	City, Country
Participant 8	City, Country
Participant 9	City, Country
Participant 10	City, Country
Participant 11	City, Country
Participant 12	City, Country
Participant 13	City, Country
Participant 14	City, Country

of material with gases. TGA is also widely used in fire research or fire engineering. It allows (a) the study of thermal behaviour of a material [2], potentially loaded with flame retardant or other fillers, or (b) the determination of kinetic parameters [3] and the extrapolation of a given material behaviour at other heating rates or at larger scales [4]. From 1986 to 2020, several round-robin studies were performed with TGA on cellulose [5,6], wood [6], synthetic polymers [7,8] or carbon fibre-reinforced thermoplastic composites [9]. The material choice was obviously related to the research domain of participants and round-robin organisers. These studies show that a good reproducibility of TGA measurements can be achieved but it strongly depends on the nature of the material. This assessment is valid for a wide range of equipment. However, these comparisons are made on reduced kinetic parameters after data processing and not on raw data. As a consequence, the assessment is less on the TGA experimental process itself than on the whole chain (measurement followed by specific post-processing).

A round robin test on thermogravimetric analysis was therefore organised to validate experimental methods and improve result quality. It gathers enough participants active in fire research to assess the reproducibility of measurements as well as to trigger discussions on best practice. Eighteen participants were initially interested in the project; fourteen sets of measurement were received provided by thirteen entities. One of the participants performed measurements with two types of equipment. Two laboratories were working together and provided only one set of measurements. The list of participants is available in Table 1.

The paper falls into four main parts after this introduction. Section 2 presents the various TGA equipments used, test material used in this round-robin, and the common protocol used by all participants together with the specificities of each participant. Then the results are presented in section 3: temperature programme conformity, mass loss curves, decomposition temperatures. Outlier detection using Grubbs method and then an alternative method is presented in section 4, before some recommendations, which falls into two parts: good practice for TGA and methods for outliers detection in TGA. Section 5 is the conclusion.

2 Experimental protocol

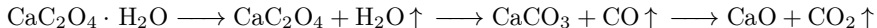
2.1 Equipment

There are several manufacturers of TGA apparatus, but all are based on the same principle: the sample is put in a crucible, which weight is measured continuously along the experiment. The complete system is put in an enclosure heated by an electric resistance, called oven, and swept by a gas flow. The aim is that the actual sample temperature T_s follows precisely a temperature programme as function of time, $T_p(t)$. This is done by using thermocouples closely connected to the sample, and using an active control of power delivered to the heating resistance as function of $T_p(t)$ and measured temperature at the thermocouple T_c . Since there is no actual way of knowing the actual sample temperature, the calibration procedure on reference material is of crucial importance. Generally it uses material with a phase change (either melting point or Curie point) in the temperature range of interest to get insight on actual sample temperature. Similarly, the systems used for weight measurements vary. They can be based on a horizontal arm, top-deck scale or a suspension load cell. Each TGA apparatus follows the above principles, but vary in the way they are designed and built. In the present work, a total of five different brands and ten different TGA apparatus were used. Participants were encouraged to provide as much information as possible. Table 2 summarises the technical specifications of these instruments as provided by the manufacturers.

For the instruments of most participants, mass calibration is performed using an automatic program realised internally. A participant verifies the auto-calibration by performing a test on 15 mg to 20 mg of CaCO_3 in air atmosphere at 20 K min^{-1} from 773 K to 1173 K. Four participants verify their weighting scale with external known masses. As already stated, temperature calibration is critical to ensure a good conformity to temperature programme. Table 3 summarises information regarding the temperature calibration of instruments; the frequency of calibration was provided by most participants although not initially requested in the round-robin procedure.

2.2 Material sample

The material used for the sample was determined to have several qualities: (1) being well studied and with stable manufacturing conditions; (2) having several decomposition steps, well separated in temperature; (3) having known values for residual mass at each step; (4) having a large amount of residue after final step (this last property was chosen to add an extra criterion to assess reproducibility). Thus, a widely known material was chosen for this study: calcium oxalate monohydrate in powder form from Sigma-Aldrich (ref. C0350000, CAS No. 5794-28-5). All the samples used in this study came from the same production batch acquired and dispatched to all the participants. Indeed, the following sequence is commonly reported for the decomposition of calcium oxalate monohydrate into calcium oxalate, then into calcium carbonate and finally into calcium oxide:



For an initial mass of sample m_0 , and m_i residual mass after step i , we define nor-

Table 2 Characteristics of the instruments

	Brand & Model	Load method	Mass sensitivity	Position and number of thermocouple
A	Netzsch STA 449 F3 Jupiter	Top	0.1 μ g	1 at the bottom of each crucible
B	Mettler Toledo SDTA 851e	Horizontal	n/a	n/a
C	Netzsch Libra	Top	0.1 μ g	Furnace and sample holder
D	Netzsch STA 449 F3 Jupiter	Top	0.1 μ g	1 at the bottom of each crucible
E	Perkin-Elmer Pyris 1	Suspension	0.1 μ g	1 at the bottom of the furnace
F	Setaram Setsys Evolution	Suspension	10 μ g	1 at the bottom of each crucible
G	Mettler Toledo TGA/DSC 1	Horizontal	5 μ g	3 at the bottom of each crucible
H	Netzsch TG 209 F1 Iris	Top	0.1 μ g	1 under the sample
I	Mettler Toledo TGA/DSC 1	Horizontal	5 μ g	3 at the bottom of each crucible
J	Netzsch STA 449 F3 Jupiter	Top	0.1 μ g	1 at the bottom of each crucible
K	Netzsch STA 449 F3 Jupiter	Top	0.1 μ g	1 at the bottom of each crucible
L	TA Instruments TGA Q500	Suspension	0.1 μ g	2 above the sample
M	Mettler Toledo TGA/DSC 2 LF/HT	Horizontal	0.1 μ g	3 at the bottom of each crucible
N	Netzsch STA 449 F1 Jupiter	Top	0.025 μ g	1 at the bottom of each crucible

Table 3 Temperature calibration of instruments; elements are denoted by their symbol in periodic classification and alloys by their commercial names; n/a: non available; (*) number of calibration points depends on crucible; last calibration for participant D was made 7 years before the tests.

	Method	Calib. points	Material	Range (K)	Frequency
A	Melting point	5	In, Sn, Bi, Zc, Al	[429:933]	n/a
B	Melting point	4	In, Sn, Zc, Al	[429:933]	1 year
C	Melting point	5	n/a	n/a	0.5 year
D	Melting point	n/a	n/a	n/a	n/a - see caption
E	Curie point	4	Fe, Ni, AlumeI Perkalloy	[303:1223]	1 year
F	Melting point	n/a	n/a	[303:1373]	1 year
G	Melting point	3	In, Al, Au	[373:1373]	1 year
H	c-DTA method	n/a	n/a	[423:1233]	n/a
I	Melting point	2	In, Al	[429:933]	1 year
J	Melting point	n/a	n/a	[429:1723]	0.25 year
K	Melting point	4 to 6(*)	n/a	[429:1337]	n/a
L	Curie point	2	n/a	[631:933]	1 month
M	Melting point	2	In, Zc	n/a	1 year
N	Melting point	6 to 8(*)	In, Sn, Bi, Zc, Al, Ag	[342:1235]	0.5 to 1 year

malised mass losses for each step as

$$ML_1 = \frac{m_0 - m_1}{m_0} \quad , \quad ML_2 = \frac{m_1 - m_2}{m_0} \quad , \quad ML_3 = \frac{m_2 - m_3}{m_0} \quad (1)$$

Using molecular masses, one may compute their theoretical values:

$$ML_1^T = 12.3\% \quad , \quad ML_2^T = 19.2\% \quad , \quad ML_3^T = 30.1\%$$

They are consistent with data in the literature [10].

2.3 Experimental protocol

One purpose of the present work is the assessment of relative scatter in data due to the different equipment. The experimental procedure was therefore as detailed as possible in order to allow as many different apparatus while still providing guidelines. When needed, this experimental procedure took into account the technological differences of the aforementioned testing equipment. Furthermore, each participant also relied on internal quality procedures adapted to its own context.

The sweeping gas had to be inert with regards to calcium oxalate and its derivatives. All participants used nitrogen except one, who employed argon instead (see Table 4). The actual value of flow rate of inert gas is dependent on the equipment and was therefore not prescribed in the experimental protocol. **This flow rate can sometimes affect the measurement [?,11]. The effect of the flow rate of the inert gas on the experimental mass loss curves for the thermal dehydration of calcium oxalate monohydrate for a horizontally supported thermogravimetric analyser has been studied in [12]. It is indicated that, for a reversible reaction, an increase in the flow rate would promote the removal of the gaseous products, and therefore, would accelerate the reaction shifting its temperature range to lower temperatures. It is recommended to maintain a sufficiently fast flow rate of a purge gas in order to maximise the efficiency of the removal of gaseous products. As a consequence participants had to ensure that this parameter did not influence their measurements as specified in [12].** At the beginning of each test, the whole experimental set-up had to be purged with the chosen inert gas for 1200 s to 1800 s before starting the actual heating program [12], to ensure that no oxygen was present in the oven when experiments started without purge.

Concerning initial mass of sample, standards recommend sample mass values over 10 mg with a maximum specified value of 100 mg [13] and 30 mg [14]. However, this recommendation is highly discussed in the literature. For a heating rate of 10 K min⁻¹, [2] mentions that a smaller sample, of size of 5 mg to 10 mg, is needed to ensure that it is in thermal equilibrium with the apparatus. These authors indeed recommend to keep the sample size as small as possible within the limits of sensibility of the apparatus which usually leads to around 5 mg. A rule of thumb indicated in [12] recommends that the sample mass times heating rate should not exceed 100 mg K min⁻¹. This value, which corresponds to 10 mg at 10 K min⁻¹ was obtained from a Biot number calculation [12]. This assumption might however not be entirely verified as suggested by several publications where no convergence is observed for PP even for sample masses as small as 0.14 mg [15], 0.83 mg (at 15 K min⁻¹) [16] or 0.11 mg (at 40 K min⁻¹) [5]. From the above considerations, sample initial mass

Table 4 Inert gas and flow rates and nature and characteristics of crucibles; note that for some participants, two flow-rates are reported: the main flow-rate plus an additional inert gas flow-rate for weighing scale protection.

	Gas	Flow rate (mL min ⁻¹)	Crucible material	Diameter (mm)	Height (mm)	Volume (μ L)	Mass (mg)
A	Nitrogen	50	Alumina	6.8	3.9	n/a	n/a
B	Nitrogen	50	Alumina	6	4.6	70	180.2
C	Nitrogen	40	Alumina	6.8	4	85	140
D	Nitrogen	30	Alumina	n/a	n/a	900	1282
E	Nitrogen	20	Ceramic	7	2.5	20	140
F	Nitrogen	40	Alumina	5	25	300	450
G	Nitrogen	20+80	Alumina	6	4.5	70	181
H	Nitrogen	20+10	Alumina	6.8	n/a	85	n/a
I	Nitrogen	80	Alumina	5	4.5	70	172
J	Argon	100	Alumina	6	4	100	152.7
K	Nitrogen	40	Alumina	6.8	4	85	147.5
L	Nitrogen	60+40	Alumina	10	4	250	632.8
M	Nitrogen	40	Alumina	n/a	n/a	n/a	n/a
N	Nitrogen	20+30	Alumina	5.9	3.7	40	138

of 5 mg and heating rate of 10 K min⁻¹ were demanded, with temperatures ranging from 300 K to 1180 K. The protocol also asked for 300 s isotherms at the beginning and end of each test. All participants complied with the 10 K min⁻¹ heating rate. They also complied with the 5 mg initial sample mass (5.12 mg $\pm$ 0.19 mg) except for participant **C** which started the experiment with a mass of 20.45 mg.

Crucibles were asked to be made of alumina, with no indication for their sizes. The sizes were not imposed because they are strongly dependent on the equipment characteristics. Crucible size may nevertheless influence the measurement. Indeed, it is better for a given sample mass to compare measurements performed on the same thickness of material to have similar diffusion rates of volatile decomposition products [12]. Crucibles were indeed composed of alumina for all participants except one participant who also used a ceramic but without specifying its nature (see Table 4). **It is interesting to note that participant D used a 1282 mg crucible, which is almost 10 times heavier than the crucible used by some of the other participants.**

No conditioning was requested for the sample in the experimental protocol, because it was thought it is dependant on each laboratory quality procedures, conditioning equipment and experience. Seven of the participants reported nevertheless to have carried out conditioning of the sample before performing measurements (see Table 5). Participants were invited to repeat each test three times but it was not mandatory. Three repeats were performed for nine sets of data.

3 Experimental results

3.1 Temperature programme conformity

First, the temperature programme conformity is analysed. All participants reported sample temperature T_s as a standard output of their apparatus. As presented in Table 2, the number and location of thermocouples depend on the apparatus model, and

Table 5 Conditioning of calcium oxalate monohydrate before TGA measurements

	Duration	Temperature (K)	Relative humidity (%)
A	Minimum of 48 h in desiccator with drierite	n/a	n/a
B	n/a	n/a	n/a
C	12 h	353	n/a
D	n/a	n/a	n/a
E	n/a	n/a	n/a
F	n/a	n/a	n/a
G	1800 s in TGA oven	300	n/a
H	Minimum of 48 h	269.15	50
I	Minimum of 48 h in desiccator with drierite	n/a	n/a
J	n/a	n/a	n/a
K	Minimum of 48 h	n/a	n/a
L	n/a	n/a	n/a
M	n/a	n/a	n/a
N	Minimum of 48 h	291	15

the critical assessment of whether T_s is actually representative of the sample temperature is not possible. However, it is significant to use these data, as they are designed to follow exactly the temperature programme as function of time $T_p(t) = \beta_p t$ during the heating phase. In order to analyse programme conformity, the estimated sample heating rate was therefore computed

$$\beta_s = \frac{dT_s}{dt} \quad (2)$$

and compared with $\beta_p = 10$ K/min. For simplicity of reading, β_s is plotted as function of T_s in Fig. 1. Note that taking derivative always generates a risk of increasing the noise because of over-sampling or under-resolved data. Therefore the sampling rate of dataset was first decreased to insure that temperature differences were actually significant. After this stage, only one dataset (dataset D) was still under-resolved and therefore noisy. Bernstein polynomials smoothing was applied to dataset D (see [17,18] for general considerations on the method and [19] for an example of its application to fire science). **This smoothing method is particularly effective for high frequency noise, such as that observed for this participant. It is therefore considered that the smoothing does not affect the data set and its use in the rest of this study.**

The curves in Fig. 1 are typical of active control with devices as Proportional-Integral-Derivative (PID) controllers with, in most cases, a transient over- or under-shoot and in one case an oscillatory behaviour. Results can be sorted in three categories: (1) participants with an over or undershoot which is smaller than 3 K/min above 350 K and smaller than 1 K/min above 400 K, hence following very well the temperature programme in the range of temperatures of interest; (2) participants with a stronger overshoot, which disappear around 500 K; (3) one participant with an oscillatory behaviour slowly damped but still present at 800 K. These facts have to be interpreted with care: since the mass loss data are plotted as function of temperature, the first order effect of non conformity in temperature programme is excluded, and the remaining effect is second order: transient heat storage in the samples and the crucibles. Therefore, it is not possible from these data to decide

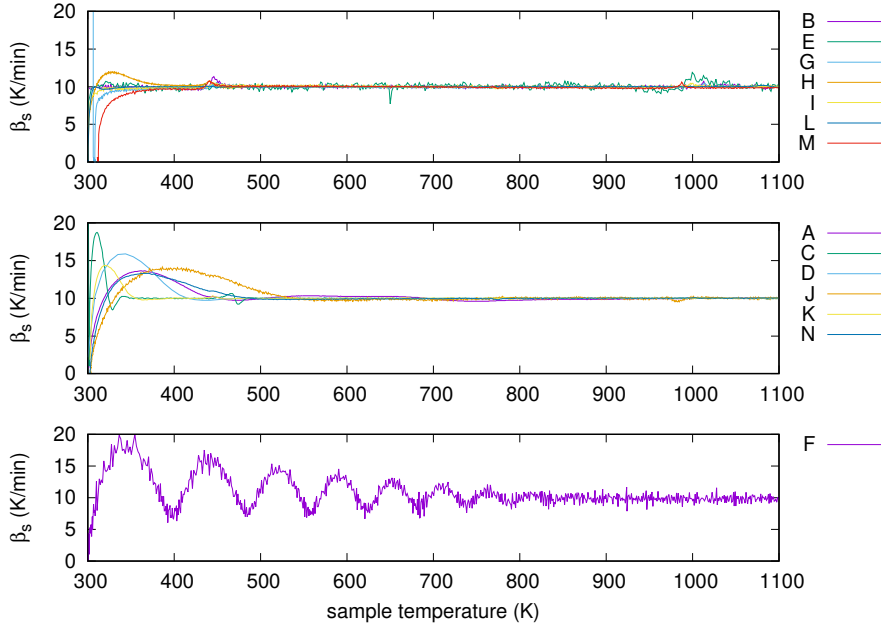


Fig. 1 Temperature programme conformity: β_s plotted as function of sample temperature T_s ; top panel: participants with conformity achieved below 400 K (see text); middle panel: participants with non oscillatory behaviour, but conformity reached slowly (note that a smoothing had to be practised for participant D, see text); bottom panel: participant with oscillatory behaviour.

whether a measurement should be excluded or not. A thorough analysis of mass loss curves is needed to go further.

3.2 Mass loss data

As stated previously, the majority of participants performed several repeats. Except for participant F, it appears visually that an excellent repeatability is reached (not shown here as curves collapse perfectly). The only exception is participant F who had a real difficulty of repeatability for step 1 (see Table 6). For the rest of the paper, only repeat 2 was kept for participant F.

Normalised masses (i.e. $m(T)/m_0$ for each test) are plotted in Fig. 2 as functions of sample temperature for the fourteen sets of data. A general tendency is clearly observable but discrepancies are visible for mass losses as well as for onset or endset temperatures. These discrepancies tend to accumulate one decomposition step after the other, to be finally much more visible for the third decomposition step.

The participants determined from their measurements the mass lost during the three decomposition steps. For data sets with repeats, an average value was first calculated for each participant and this value later used in the determination of the overall average. Normalised mass losses are given in Table 6.

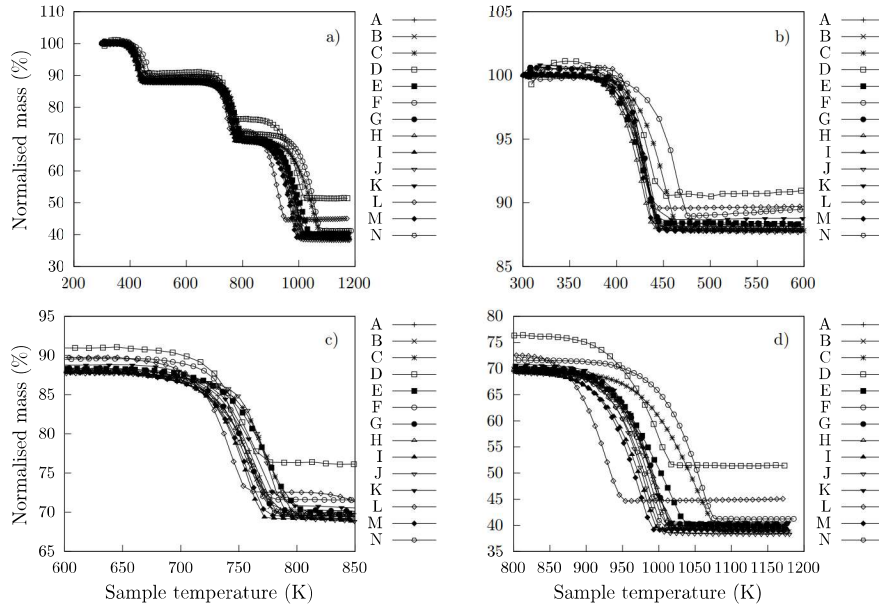


Fig. 2 Normalised mass as a function of sample temperature – Comparison between participants including all dataset

For each decomposition step, the average normalised mass loss is lower than the theoretical normalised mass loss. This means that at each step the normalised residual mass is larger than its theoretical value. This might be caused by incomplete reaction, however this is not likely from the data since the mass residuals appears to be well stabilised. The most probable explanation would be the presence of additive in the samples. However, since all samples were taken from the same batch of production, this does not raise concern as far as this round robin study is concerned. Furthermore, as expected for increasing decomposition steps, it is observed an increase in the standard deviation as well as in the difference between the maximal and minimal measurements. This can be seen as a build-up of measurement errors from previous decomposition steps.

3.3 Decomposition temperatures

Participants had to determine from their measurements the onset, peak and end-set temperatures for each decomposition step, using their own methods. For data sets with repeats, an average value was first calculated and later used in the determination of an overall average. Unlike for the mass lost during a decomposition step (see section 3.2), no theoretical value exists as a reference for the decomposition temperature, which depends on the experimental procedure (*e.g.* initial sample mass or heating rate).

Table 6 Normalised mass loss ML_1 , ML_2 and ML_3 (in %) as defined by eq. (1); note that the largest uncertainty in mass measurement for F is $10\mu\text{g}$ representing 0.2 % of initial mass; uncertainties for G and I is 0.2 % of initial mass, and less than 0.002 % of initial mass for A, C, D, H, J, K, L, M and N (see Table 2). R-1 to R-3 correspond to the test replicates, if applicable.

	ML_1	ML_2	ML_3
A	12.8	18.0	30.0
B	11.8	18.1	29.9
C	12.2	18.7	30.3
D	9.3	12.6	20.8
E-R1	11.7	18.6	29.9
E-R2	11.7	18.6	29.8
E-R3	11.8	18.9	30.0
F-R1	11.1	18.3	29.9
F-R2	10.7	18.0	30.2
F-R3	10.1	18.0	31.3
G-R1	12.2	18.2	29.9
G-R2	12.3	18.1	29.9
G-R3	12.2	18.2	29.7
H	12.1	18.4	30.3
I-R1	12.0	19.0	30.0
I-R2	12.0	19.0	30.0
I-R3	12.0	19.0	30.0
J-R1	12.3	18.7	30.6
J-R2	12.4	18.5	30.2
J-R3	12.3	18.5	30.3
K-R1	11.2	18.3	29.9
K-R2	11.2	18.3	29.9
K-R3	11.2	17.9	30.1
L	11.0	17.2	27.9
M	12.0	18.0	30.0
N-R1	12.1	18.5	29.7
N-R2	12.1	18.4	30.0
N-R3	12.1	18.5	30.1
Theory [10]	12.3	19.2	30.1
Median	12.0	18.2	30.0
Average	11.6	17.9	29.3
Standard deviation	0.8	1.5	2.4

Data are also plotted in Fig. 3 where temperatures reported for the 1st step are in black with no fill, cross-hatched for the 2nd step and hatched for the 3rd step. The bottom line of each rectangle corresponds to the onset temperature, the centre line to the peak temperature and the top line to the end-set temperature. These temperatures are also compared to the averaged onset, peak and end-set temperatures plotted with their standard deviations on the right hand side of the figure.

To analyse these results, a comparison is made between measurements from participants A and G which are very close (see Fig. 2). The lengths of the three rectangles from participant A in Fig. 3 are however much bigger than those of participant G. For the first, second and third decomposition steps, the rectangle lengths for participant A are respectively 52 K, 82 K and 110 K, whereas for participant G these rectangle lengths are respectively 32 K, 47 K and 60 K. Both participants did not use the same standard or protocol to determine the onset and end-set temperatures which led to such differences. On the contrary, the reported peak temperatures of these partic-

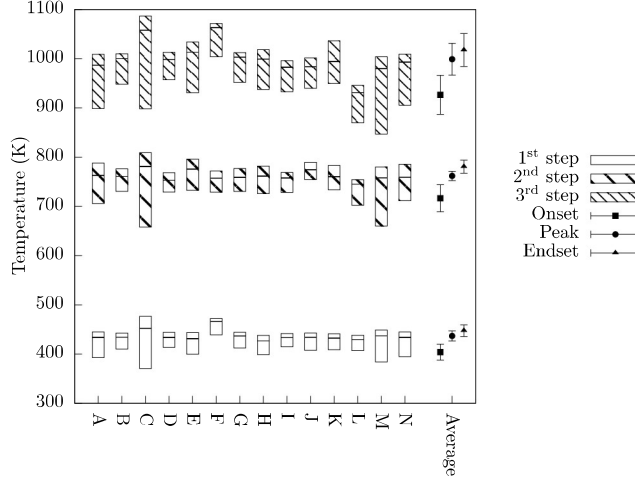


Fig. 3 Onset, peak and endset temperatures for the three decomposition steps compared with average values

ipants where much closer: an absolute difference of 2 K for the first decomposition step between peak temperatures of participants A and G, 4 K for the second decomposition step and 16 K for the third decomposition step. The determination of peak temperature appears to be less dependent on the chosen standard or method. Finally, as observed previously in section 3.2, the standard deviations reported in Table 6 increase with increasing decomposition steps due to an accumulation of measurement errors.

4 Discussion

4.1 Detection of outliers: Grubbs test method

4.1.1 Systematic tests for outlier detection

From previous section, it is clear that a group of results are well concentrated while other are further away and can be seen as outliers. The aim of the present section is to apply a generic method for outlier detection and see whether general conclusions can be drawn from the analysis. The method chosen for outlier detection is Grubbs test [20,21] **which is generally recommended for small data sets (*i.e.* less than 21 test results)**. The principle of Grubbs test for a series of data is the following.

Starting with n data, $x_1, x_2, \dots, x_n$ sorted in increasing order. First, compute V_n , the variance of the full set of data, and V_{n-1} the variance of data excluding the largest one. Assuming data follow Gaussian pdf, statistical theory shows that V_{n-1}/V_n has to be close to one with a known probability. Hence, for outlier detection, it is decided to exclude or not the largest data depending on the value of V_{n-1}/V_n compared with

Table 7 Statistics of normalised mass loss ML_1 , ML_2 and ML_3 (in %) as defined by eq. (1) excluding outliers D, F, K and L.

	ML_1	ML_2	ML_3
Theory [10]	12.3	19.2	30.1
Median	12.1	18.45	30.0
Average	12.0	18.4	30.1
Standard deviation	0.2	0.4	0.2

a tabulated value depending on the number of data and the significance threshold retained. The procedure is then carried out similarly for the smallest of the data, and so on iteratively until no outlier is found. One of the known dangers of Grubbs test is over-exclusion, since variance is recalculated at each step after having excluded a new outlier; thus, an overly stringent confidence level should not be used.

We nevertheless decided to use Grubbs test because of its generality and broad application. The confidence level retained in the present work is 2.5 %. Since the process is iterative and monovariate, we decided to work first on mass loss and then on temperature on the remaining set.

4.1.2 Mass loss data

From the mass loss data with 2.5 % confidence level, results from participants D, F, K and L are then detected as outliers, leaving us with 10 datasets: A, B, C, E, G, H, I, J, M, and N.

As mentioned previously, crucibles of participants D and L were respectively 8 times and 4 times heavier than those of other participants which may explain these observed differences in mass measurements. Indeed, as the sample mass was set at 5 mg in the experimental protocol, it is suspected that these instruments did not provide good data because the prescribed sample mass is outside their operating conditions or because of an issue with the baseline correction/conditioning step, which later propagated into these difference in apparent mass loss in each step. For participants F and K, they are considered as outliers due to a poor representation of the first decomposition step.

The average value without participant data detected as mass outliers is compared to the theory in table 7. These data can be directly compared to those listed in table 6 which contain measurements from all the participants. The exclusion of mass outliers significantly reduces the standard deviation which is lower than 0.35% for the third decomposition step compared to 2.4% with mass outliers. The difference between the maximal and minimal values is divided by up to 16 for the third decomposition step. And more importantly, the relative differences between the average and theoretical values are reduced from 5.7 %, 6.8 % and 2.7 % for respectively the 1st, 2nd and 3rd decomposition steps to 2.4 %, 4.2 % and 0 %. Normalised masses as functions of sample temperature are plotted in Fig. 4 without data from participants detected as mass outliers. The dispersion of mass measurements is smaller along the vertical axis than the one observed in Fig. 2. However, a wide dispersion is still visible along the horizontal axis, especially for participant C. which means that some data sets are more precise for mass measurements than for temperature measurements, and show the need to detect also temperature outliers.

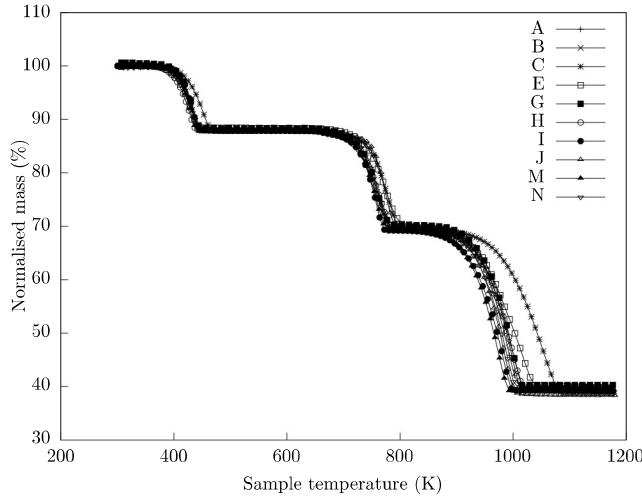


Fig. 4 Comparison between participants of the normalised mass as a function of sample temperature (without measurements having mass outliers); letters A to N refer to dataset presented on table 2.

4.1.3 Temperature data

Grubbs tests were applied to all datasets, retained the datasets which were identified as mass outliers. Each decomposition step was considered separately. All the participants provided measurements which were compliant with at least eight of the nine investigated temperatures (*i.e.* onset, peak and end-set temperatures of the three decomposition steps). Datasets F, C, H, L and M were detected consecutively as outliers in the iterative process. Figure 5 presents temperature data without participant measurements containing temperature outliers, to be compared with Fig. 3. Table 8 gathers statistics for each decomposition step. Standard deviations are drastically reduced: they are divided by two for onset temperatures, between 1.3 and 6.8 for peak temperatures and between 1.4 and 8.4 for end-set temperatures.

Mass measurements as functions of time are plotted in Fig. 6 without participant data detected as temperature outliers. Compared to Fig. 2, measurements in Fig. 6 are less scattered along the horizontal temperature axis. It is interesting to observe that data from participant D which were detected as outliers for mass measurement are not for temperature measurement. This is why this thermogram seems to be detached from other measurements in Fig. 6.

4.1.4 Combined mass and temperature outliers

As a whole, D, F, K and L are identified as mass outliers, and F, C, H, L and M as temperature outliers. F and L are double outliers. We are left with 7 non-outliers datasets: A, B, E, G, I, J and N. Figure 7 gives mass loss data from these remaining datasets and Table 9 the associated statistics.

Table 8 Onset, peak and end-set temperatures (expressed in K) after outliers exclusion (see text) for each decomposition step; T_i^X is for step i with $X = O$ for Onset, $X = P$ for Peak, and $X = E$ for end-set.

	T_1^O	T_1^P	T_1^E	T_2^O	T_2^P	T_2^E	T_3^O	T_3^P	T_3^E
Min	393.0	431.3	441.1	706.0	753.2	768.3	899.0	982.7	996.0
Average	406.1	433.8	443.4	728.6	762.6	781.4	935.1	995.2	1013.6
Max	414.7	436.7	445.0	754.4	775.7	796.0	957.4	1013.3	1036.3
Std. dev.	8.2	1.5	1.4	13.8	7.5	9.5	20.6	10.0	13.4

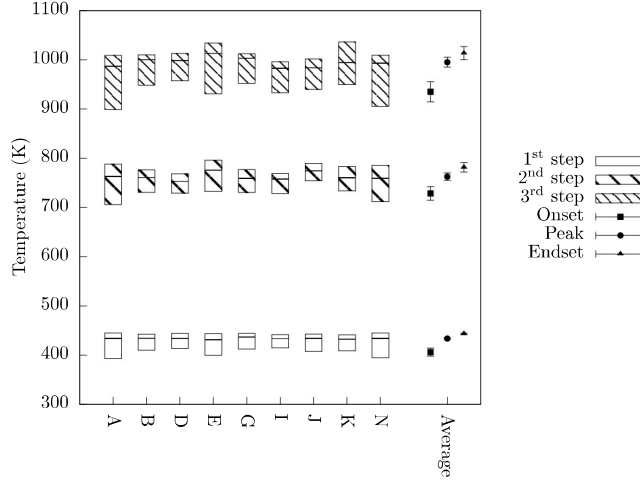


Fig. 5 Onset, peak and end-set temperatures without temperature outliers for the three decomposition steps compared with average values

Table 9 Min, Average, Max, and Standard deviation for mass loss (expressed in %), onset, peak and end-set temperatures (expressed in K) for the three decomposition steps without outliers, retaining only A, B, E, G, I, J and N.

	ML_1	ML_2	ML_3	T_1^O	T_1^P	T_1^E	T_2^O	T_2^P	T_2^E	T_3^O	T_3^P	T_3^E
Min	11.7	18.0	29.9	393	431	442	706	758	769	899	983	996
Ave.	12.1	18.4	30.1	405	434	444	728	764	783	930	995	1010
Max	12.8	19.0	30.5	414	437	445	754	776	796	952	1013	1034
St.d.	0.34	0.34	0.24	8.1	1.5	1.2	15	7	9	19	11	11

4.2 Alternative method for outlier detection

From the above outlier detection using Grubbs method, one can argue that only 7 participants (A, B, E, G, I, J, and N) out of 14 provided data correct for both mass loss and temperature. However, the method chosen in the present study to select the measurements has two potential limitations. First, Grubbs test works from brute force with a chosen confidence level (2.5 %) which might be too harsh, especially due to the iterative nature of the test which recalculated statistics at each

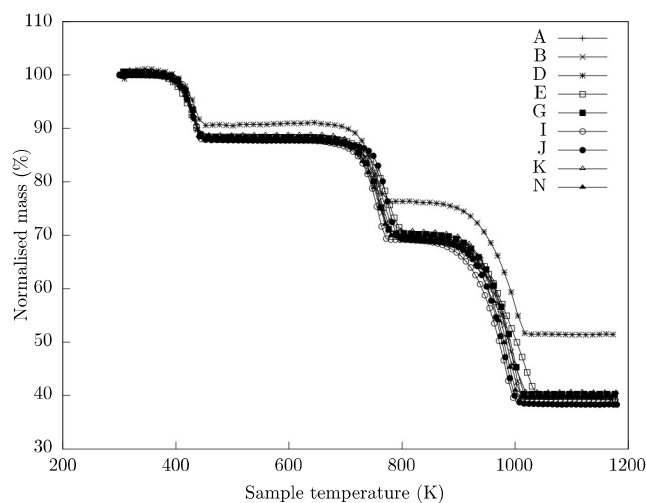


Fig. 6 Comparison between participants of the normalised mass as a function of sample temperature (without measurements having temperature outliers)

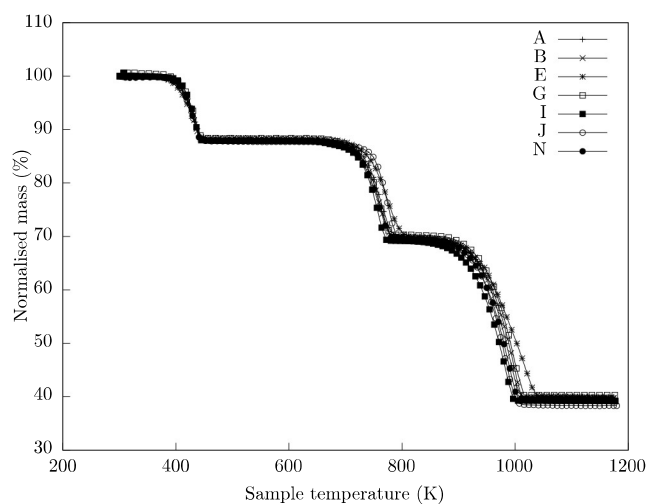


Fig. 7 Normalised mass as a function of sample temperature without participants data with identified mass and temperature outliers

step. Second, the analysis was based on mass losses and onset, peak and end-set temperatures for the three decomposition steps provided by the participant using their own estimation methods. Thus, the analysis detects non exactly measurement outliers, but measurement and processing method outliers. An incorrect estimation of any of the requested mass losses or temperatures may sanction the entire dataset even

if the measurement itself was correct. It would be a much better strategy to exclude outliers from full mass-temperature relationship. However, there is no systematic method for doing this.

We decided to use a two step method inspired from the Z-score method defined in standard ISO13528 [22]. At each step, we compute average and standard deviation as function of temperature using n selected datasets:

$$\langle m \rangle (T) = \frac{1}{n} \sum_i m_i(T) \quad (3)$$

$$\sigma(T) = \sqrt{\frac{1}{n} \sum_i (m_i(T) - \langle m \rangle (T))^2} \quad (4)$$

and finally define

$$Z_i(T) = \frac{m_i(T) - \langle m \rangle (T)}{\sigma(T)}. \quad (5)$$

We then plot the graph $Z_i(T)$ and exclude dataset which have a significant range of temperature with $|Z_i| > 2$. **This value corresponds to a 5% percentile criterion for a Gaussian variable, which is common in such studies.** For the first step, we keep all the dataset, leading to Fig. 8(a). This leads to the exclusion of datasets D, F, and K. For the second step, average and standard deviation are re-computed after the exclusion, and functions $Z_i(T)$ are computed again. This leads to Fig. 8(b). Now it can be seen that C is an outlier for the third step, while L and N are on the edge, and all the other dataset are well grouped in the middle of the range.

4.3 Recommendations

From the above analysis, it can be seen that, from a Grubbs test point of view, half of the participants would be considered as outliers. On the other hand, a more detailed analysis based on methodology inspired from Z-score would lead to exclude 4 participants out of 14. Recommendations should be grouped into two categories: recommendations of good practice when performing TGA, and recommendations for outliers detection.

4.3.1 Good practice for performing TGA analysis

From the Grubbs analysis, it was complicated to identify reasons for dataset to be excluded. On the other hand, from the second analysis, a few points may be raised.

First, going back to Fig. 1, it appears that F had a real issue of oscillating sample temperature, while C, D and K where in the middle panel showing a strong overshoot. **It is unclear whether a more frequent calibration could have brought better results. Calibration periodicity for participant K was unknown, F was one year and D was several years (see Table 3).**

Concerning the initial mass of the sample, only C was different from the other participants, and it may also explain the difference of result. This advocate to follow [12] recommendation to keep samples as small as reasonably possible. The crucible size may also be analysed. The fact that D and F used a relatively heavy

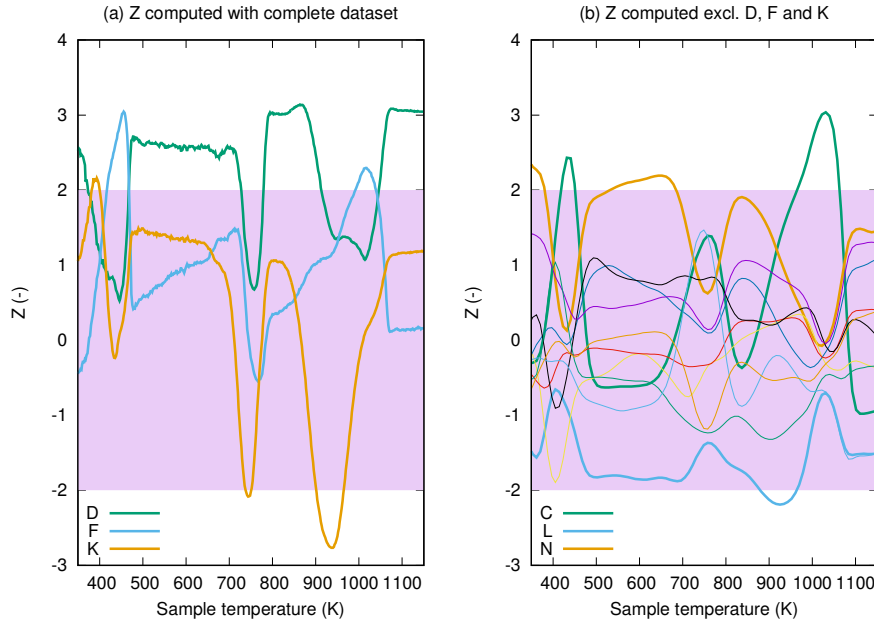


Fig. 8 Generalised Z-score (see text); (a) all dataset are included in the computation of average and standard deviation but only D, F and K are shown; (b) datasets D, F, and K are excluded in the computation of average and standard deviation; all datasets excluding D, F, and K are shown, and lines are smoothed for improved visualisation – only linestyles for C, L and N are specified.

crucible could be seen as a clue, see Table 4. However, C and K crucibles were well in the range of other tests, while L was the second heaviest, at the edge of in- and outliers. Nevertheless, it seems that the effect of crucible mass is poorly understood, and the study of the influence of inert mass in TGA may be of interest in order to investigate scattering among test results.

Finally, performing three repeats serves as a good self-verification practice. Most participants achieved very good repeatability indicating that it is possible to get consistent results from this material. However, even with an excellent repeatability, measurements can deviate from theoretical values.

4.3.2 Outlier detection

Comparing Grubbs method and the two-step method, it seems that the former may sometimes lead to a too high selectivity. In the context of round robin procedure, the main advice here seems limited to a careful inspection of data before using a systematic iterative procedure.

Now, the issue may be formulated in another manner. Suppose a given laboratory performs a TGA in conditions similar to the ones presented here. How will it be possible to analyse these new data in the light of the present ones? For this the answer is double, as described now. First, ML_1 , ML_2 , ML_3 , T_1^O , T_1^P , T_1^E , T_2^O , T_2^P , T_2^E , T_3^O , T_3^P , and T_3^E issued from the new analysis should be compared with

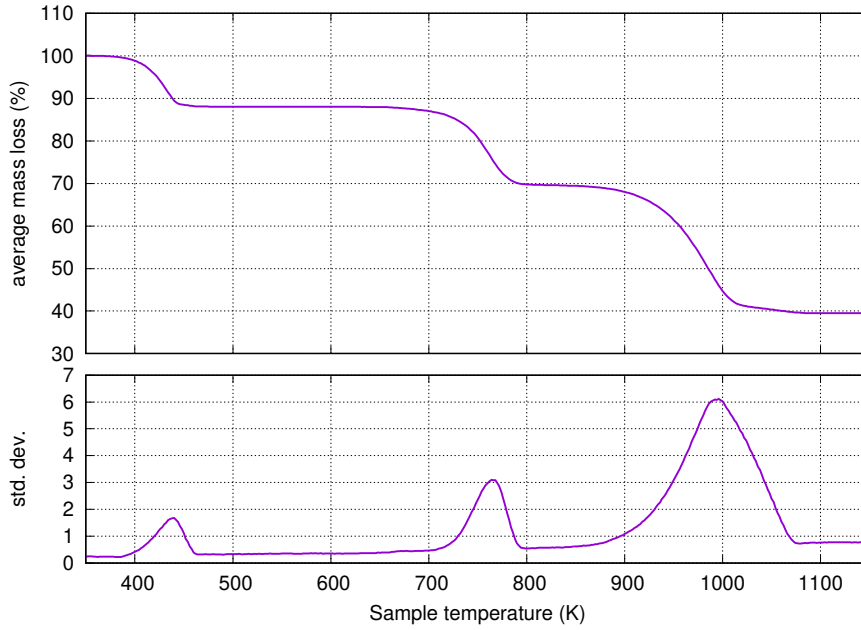


Fig. 9 Normalised mass loss curve (%) as function of sample temperature (K); top panel: average of all datasets except D, F and K; bottom panel: standard deviation of all datasets except D, F and K.

data given in Table 9. If all fall in the range given by average plus or minus two times standard deviation, the new data are deemed to be satisfactory according to common practice (see also [13]). Second, mass loss curve data should be compared with average and standard deviation excluding D, F and K as described in section 4.2. The computation of $Z(T)$ allows to decide whether the new dataset is satisfactory with regards to the alternative test. Average and standard deviation excluding D, F, and K are plotted on Fig. 9 and available on request to the corresponding author.

5 Conclusion

A round-robin test was organised on calcium oxalate monohydrate with thirteen participants with a high level of expertise in the fire domain (leading to fourteen dataset as explained in introduction). The main purpose of this work was the assessment of scatter in data due to equipment. The experimental procedure was consequently as detailed as possible even though it cannot have been entirely prescribed because of the different technologies of thermogravimetry analysers. The investigated material is frequently used to evaluate the proper functioning of a given equipment because its decomposition steps are well-documented and known to be repeatable and reproducible. During the presentation of the results, it was made apparent that temperature programme conformity is a real issue.

The main results of the paper are in section 4. First, using Grubbs method for outliers detection, statistics presented in Table 9 may be used for a stringent test of

new TGA runs with calcium oxalate and associated temperature of reaction identification algorithms. Then, in the same context, for the detection of outliers in the mass loss curves themselves, ideas of section 4.2 apply. Thus, this work confirms that a verification with calcium oxalate is a good screening method, which may be used by laboratories. Within a laboratory it can also assist in finding problems with equipment accuracy otherwise not possible to distinguish. This, together with good practice outlined in section 4, can serve as a validation dataset for the experimentalist, and should help the community to increase the validity of tests with other materials of interest for fire science.

Acknowledgements This work was supported by the International Forum of Fire Research Directors which validated the material selection and experimental procedure. The FORUM (www.fireforum.org) is a group of leaders of fire research organizations from around the world that meet on an annual basis to discuss emerging fire related research topics throughout the world. It has been active since 1991. Its aim is to reduce the burden of fire (including the loss of life and property, and effects of fire on the environment and heritage) through international cooperation on fire research. An important membership requirement is that the applicant be responsible for establishing research priorities and allocating resources for fire safety research in their parent organization. Authors are grateful for their help and recommendations.

Competing interests

The authors declare that they have not identified any conflict of interest.

References

1. Sergey Vyazovkin. Thermogravimetric Analysis. In Elton N. Kaufmann, editor, *Characterization of Materials*, page com029.pub2. John Wiley & Sons, Inc., Hoboken, NJ, USA, October 2012.
2. Artur Witkowski, Anna A. Stec, and T. Richard Hull. Thermal Decomposition of Polymeric Materials. In Morgan J. Hurley, Daniel Gottuk, John R. Hall, Kazunori Harada, Erica Kuligowski, Milosh Puchovsky, José Torero, John M. Watts, and Christopher Wieczorek, editors, *SFPE Handbook of Fire Protection Engineering*, pages 167–254. Springer New York, New York, NY, 2016.
3. Sergey Vyazovkin, Alan K. Burnham, José M. Criado, Luis A. Pérez-Maqueda, Crisan Popescu, and Nicolas Sbirrazzuoli. ICTAC Kinetics Committee recommendations for performing kinetic computations on thermal analysis data. *Thermochimica Acta*, 520(1-2):1–19, June 2011.
4. Thomas Rogaume. Thermal decomposition of solid fuels. objectives, challenges and modelling. *Journal of Physics: Conference Series*, 1107:022001, 11 2018.
5. Morten Grønli, Michael Jerry Antal, and Gábor Várhegyi. A Round-Robin Study of Cellulose Pyrolysis Kinetics by Thermogravimetry. *Ind. Eng. Chem. Res.*, 38(6):2238–2244, June 1999.
6. Andrés Anca-Couce, Christos Tsekos, Stefan Retschitzegger, Francesco Zimbardi, Axel Funke, Scott Banks, Tzouliana Kraia, Paula Marques, Robert Scharler, Wiebren de Jong, and Norbert Kienzl. Biomass pyrolysis TGA assessment with an international round robin. *Fuel*, 276:118002, September 2020.
7. Michihiko Momota, Takeo Ozawa, and Hsaaki Kanetsuna. Thermal analysis of polymer samples by a round robin method. *Thermochimica Acta*, 159:125–137, January 1990.
8. T. Ozawa. Industrial standardization of thermal analysis in Japan. *Journal of Thermal Analysis*, 40(3):1379–1398, September 1993.
9. Masahiro Funabashi, Fumi Ninomiya, Akihiro Oishi, Wataru Mizuno, Kimitaka Tahira, Shogo Tanaka, and Masao Kunioka. Round Robin Tests to Determine Fiber Content of Carbon Fiber-Reinforced Thermoplastic Composites by Combustion and Thermogravimetry. *Journal of Polymers*, 2017:1–10, 2017.

10. H.L Anderson, A Kemmler, G.W.H Höhne, K Heldt, and R Strey. Round robin test on the kinetic evaluation of a complex solid state reaction from 13 european laboratories. part 1. kinetic tg-analysis. *Thermochimica Acta*, 332(1):33–53, 1999.
11. David Lázaro, Mariano Lázaro, Alain Alonso, Pedro Lázaro, and Daniel Alvear. Influence of the STA boundary conditions on thermal decomposition of thermoplastic polymers. *J Therm Anal Calorim*, 138(4):2457–2468, November 2019.
12. Sergey Vyazovkin, Konstantinos Chrissafis, Maria Laura Di Lorenzo, Nobuyoshi Koga, Michèle Pijolat, Bertrand Roduit, Nicolas Sbirrazzuoli, and Joan Josep Suñol. ICTAC Kinetics Committee recommendations for collecting experimental thermal analysis data for kinetic computations. *Thermochimica Acta*, 590:1–23, August 2014.
13. ISO11358-1:2014. Plastics Thermogravimetry (TG) of polymers Part 1: General principles, 2014.
14. ASTM E1131-20. *Standard Test Method for Compositional Analysis by Thermogravimetry*. ASTM International, West Conshohocken, PA, 2014.
15. Aurélien Thiry-Muller. *Modélisation de la décomposition thermique des solides*. PhD thesis, Université de Lorraine, 2018.
16. Dawid Stawski. The effect of sample weight in thermogravimetric analysis of low viscosity polypropylene in air atmosphere. *Polymer Testing*, 28(2):223–225, April 2009.
17. J. Dieudonné. *Calcul infinitésimal*. Hermann, 1968.
18. A. Blejec. Bézier curves: simple smoothers of noisy data. In A. Ferligoj and A. Kramberger, editors, *Developments in statistics and methodology*. 1999.
19. A. Suzanne and P. Carlotti. Heat release rate of burning luggage in trains. In M. Tabarra, editor, *International Symposium on Aerodynamics and Fire in Vehicle Tunnels (ISAVFT)*. 2022.
20. Frank E. Grubbs. Sample criteria for testing outlying observations. *Annals of Mathematical Statistics*, 21(1):27–58, 1950.
21. Frank E. Grubbs and Glenn Beck. Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14(4):847–854, 1972.
22. ISO13528. Statistical methods for use in proficiency testing by interlaboratory comparison, 2015.