Available online at www.sciencedirect.com
SCIENCE DIRECT®
Thermochimica Acta 406 (2003) 177–183
thermochimica acta
www.elsevier.com/locate/tca

Local polynomial estimation of TGA derivatives using logistic regression for pilot bandwidth selection

Salvador Naya [a], Ricardo Cao [a], Ramón Artiaga [b,*]

[a] Department of Mathematics, Spain Facultad de Informática, Campus Elviña, University of A Coruña, A Coruña, Spain
[b] Department of Industrial Engineering II, Spain Escuela Politécnica Superior Esteiro, University of A Coruña, 15403 Ferrol, Spain

Received 9 August 2002; received in revised form 7 April 2003; accepted 17 April 2003

Abstract

A method is proposed to find a function that fits traces resulting from the decomposition of polymers during a thermogravimetric analysis (TGA) experiment.

The aim of this work is to find an expression that relates the weight of the sample with time (or temperature) and to study certain features like derivatives and critical points. It has been possible to keep in the asymptotic trend of the TGA trace by using a local logistic type regression. The typical asymptotic tendency of the weight loss is perfectly reproduced.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Bandwidth; Kernel function; Logistic regression; TGA curve

1. Introduction

Although each polymer may show a specific behavior when subjected to temperatures that imply degradation, all thermogravimetric analysis (TGA) traces obtained from the analysis of polymers in the range of thermal decomposition have two features: (1) the weight continuously decreases with time (and temperature) and (2) each degradative process shows, at both ends, an asymptotic trend.

It has been assumed that time (or temperature), denoted by X, and the weight of the sample, denoted by Y, is related by the following expression:

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad E(\varepsilon_i) = 0,$$

so m is the regression function for Y given X and ε_i is the error.

One of the most frequently used parametric models for fitting regression curves are polynomial. However, the shape of the TGA traces, showing horizontal asymptotes at the beginning and the end of the reaction, make the polynomial fitting not suitable. This problem can be overcome by the use of logistic functions that preserve the asymptotic trend of the TGA traces.

2. Logistic regression

Since the TGA curve takes values, Y_i, between some minimum (A) and some maximum (B), the expression of the standard logistic regression model (where the value of the dependent variable is usually 0 or 1) should be rescaled in accordance with the following

* Corresponding author. Fax: +34-981-337410.
E-mail address: rartiaga@udc.es (R. Artiaga).

0040-6031/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0040-6031(03)00261-2

expression:

$$y = A + \frac{(B - A)\exp(P_k(x))}{1 + \exp(P_k(x))} = \frac{A + B\exp(P_k(x))}{1 + \exp(P_k(x))}, \tag{1}$$

where $P_k(x)$ is a kth degree polynomial, $P_k(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_k x^k$. The parameters of the model can be easily estimated by transforming Eq. (1) in the following form:

$$y(1 + \exp(P_k(x))) = A + B\exp(P_k(x)) \Leftrightarrow P_k(x) = \ln\left(\frac{y - A}{B - y}\right).$$

We estimate the coefficients of $P_k(x)$ by least squares from the transformed data points: the horizontal coordinates do change (the reaction time or temperature) while the new vertical coordinates are $\ln((y - A)/(B - y))$, where y is the original one.

Once the polynomial has been estimated it is used in Eq. (1) to find the estimated model. This kind of parametric fitting could work well in some TGA traces. In that case, this parametric fitting may serve as final estimation for the regression function. As an example of this fitting, a single step degradation, extracted from a thermogravimetric analysis of commercial PVC, is shown in Fig. 1. A sample of 13 mg was subjected to a heating ramp from 25 to 600 °C at 10 K/min in a Rheometric STA 1500. Open aluminum crucibles were used and an argon flow of 50 ml/min was kept along the experiment.

Fig. 1 shows the TGA trace and the parametric fit, with a polynomial of fifth degree. Although the experiment was larger, the area of interest ranges from the beginning of the experiment to 360 °C. Since the both sets of data match, they were plotted separately. Fig. 2 shows the first derivative obtained from the row data by using the RSI Orchestrator, thermal analysis software that is based in the Taylor series expansion, and the first derivative of the parametric fit.

3. Local polynomial nonparametric regression

The analysis by TGA of many materials results in more or less complex traces that do not allow a simple parametric fit like the one previously described, since although it reproduces the asymptoticity at the beginning and end of the reaction, it takes into account for one single step process. However, it is possible to use a local regression polynomial model instead.

These models are based in the idea of smoothing the points cloud by locally fitting to a polynomial for estimating the value of $m(x)$, at each point [1]. If a p-degree polynomial is taken (in powers of $X_i - x$) the model is

$$Y_i = \sum_{j=0}^{p} \beta_j (X_i - x)^j + \varepsilon_i,$$

for the X_i in a neighborhood of the x point, whose amplitude is denoted by h, the bandwidth or smoothing parameter. Using the notation

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right),$$

where K is the kernel function (typically a nonnegative symmetric function that integrates out to 1), we will have the expression:

$$\psi(\vec{\beta}, x, X_1, X_2, \ldots, X_n, Y_n) = \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (X_j - x)^j \right\}^2 K_h(X_i - x).$$

It is possible to estimate the regression function and its derivatives at the x point by using the $\hat{\beta}_j$ values that minimize the previous weighted squares sum:

$$\hat{m}^{(j)}(x) = j!\hat{\beta}_j, \quad j = 0, 1, \ldots, p.$$

For the lineal case, it turns out:

$$\hat{m}(x) = \hat{\beta}_0(x), \quad \hat{m}'(x) = \hat{\beta}_1(x),$$

4. Generalized logistic local regression model

One problem, in order to estimate the curve, is the choice of the optimal bandwidth parameter. The method proposed by Ruppert et al. [2] is used for this purpose in local polynomial regression. Let us assume that X has a fixed design roughly equally spaced in $[0, T]$. By a plug-in method, these authors estimate the asymptotically optimal bandwidth (in the sense of

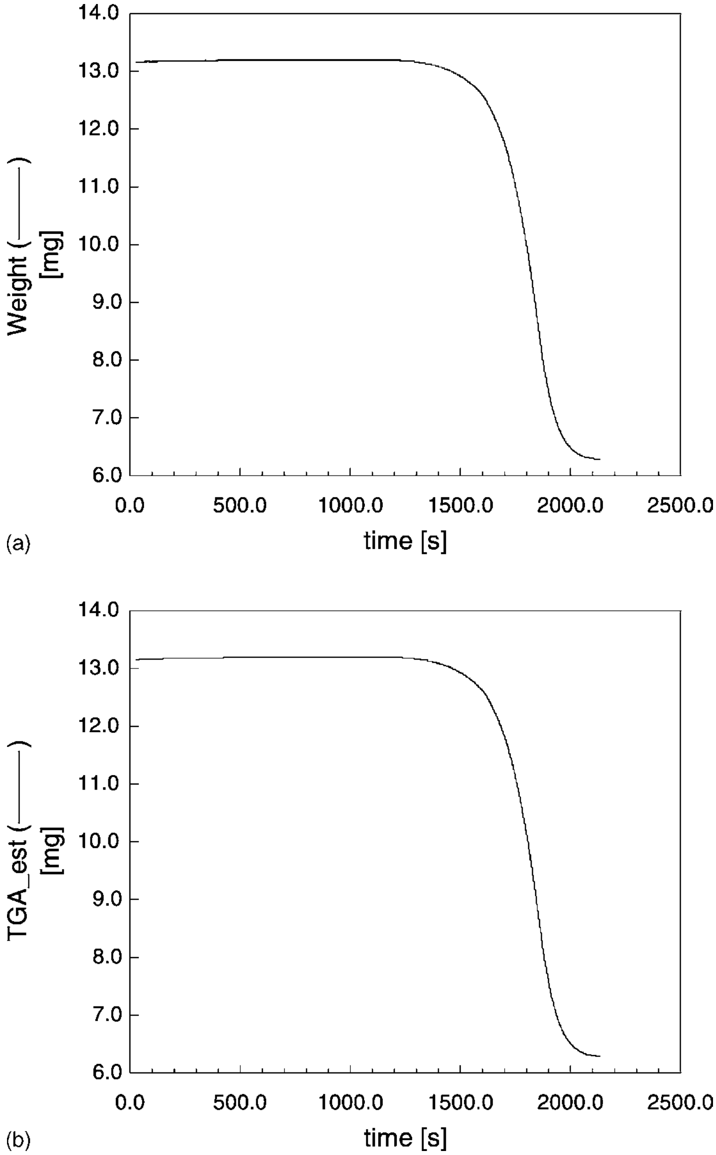


Fig. 1. The TGA trace (a) and the fifth degree polynomial logistic regression estimation (b) of the TGA trace of a commercial PVC.

MISE) for estimating the υth derivative of the regression function,

$$h_{\mathrm{opt}} = C_{\upsilon,p}(K)\left(\frac{\sigma^2 T}{n\int (m^{(p+1)}(x))^2\,\mathrm{d}x}\right)^{1/(2p+3)}, \quad (2)$$

when a p-degree local polynomial estimator is used, where $C_{\upsilon,p}(K)$ is a constant whose value has been reported for the different kinds of kernel functions K,

$$C_{\upsilon,p}(K) = \left(\frac{((p+1)!)^2(2\upsilon+1)\int (K_\upsilon^*(t))^2\,\mathrm{d}t}{2(p+1-\upsilon)\{\int t^{p+1}K_\upsilon^*(t)\,\mathrm{d}t\}^2}\right)^{1/(2p+3)},$$

$$K_\upsilon^*(t) = \left(\sum_{l=0}^{p} S^{\upsilon l}t^l\right)K(t),$$

where $S^{\upsilon l}$ are the elements of the S^{-1} matrix, where $S = (\mu_{j+l})_{j,l=0}^{p}$ with $\mu_j = \int u^j K(u)\,\mathrm{d}u$ and σ^2 is the variance of ε_i.

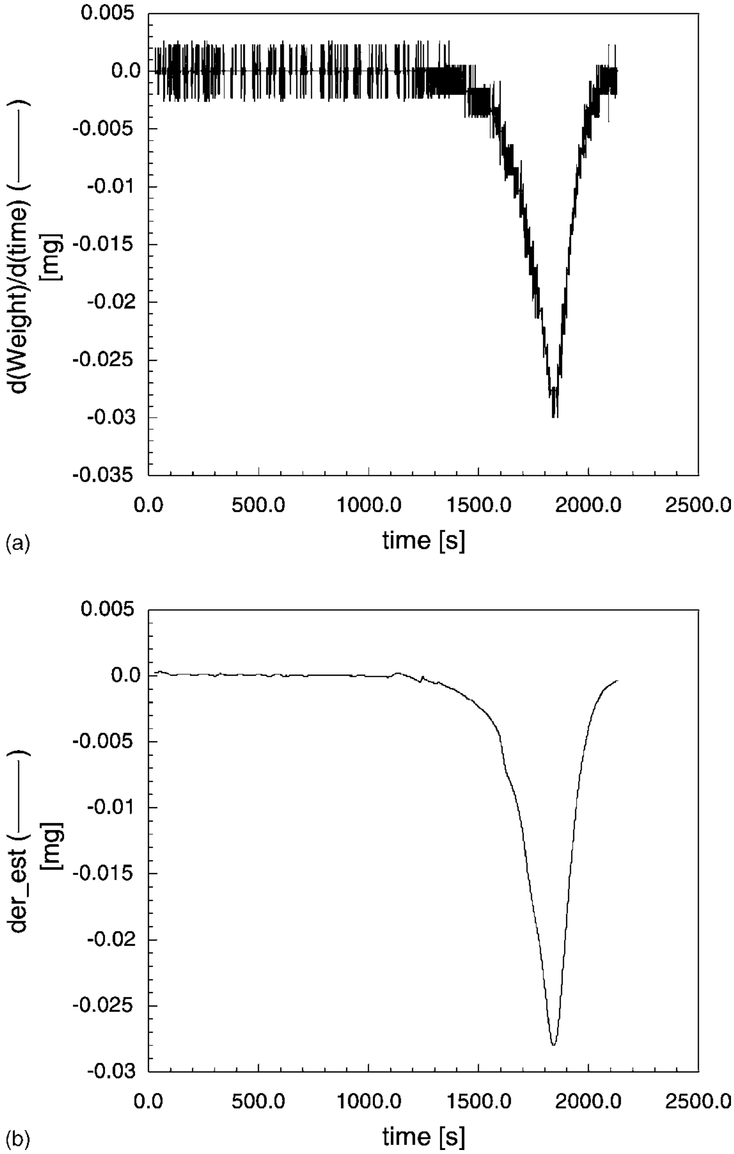


Fig. 2. The DTG obtained by the Orchestrator software (a) and the fifth degree polynomial logistic regression estimation of the DTG (b) in the case of PVC.

The first difficulty consists in estimating the denominator of expression (2) that includes the $(p+1)$-derivative of the regression function that we are estimating. To overcome this problem, we started from a parametric model for this function and then, the proposed function was derived $(p+1)$ times. The derivative in expression (2) is then replaced by its parametric estimation.

We propose a logistic model (1) for estimating the integral that appears in the denominator of the optimal bandwidth, where a $k = p + 3$ degree polynomial has been used.

When using a nonparametric model for the estimation of $\int (m^{(p+1)}(x))^2\,\mathrm{d}x$, the errors obtained in the estimation of the curve and in the estimation of its derivatives are typically smaller

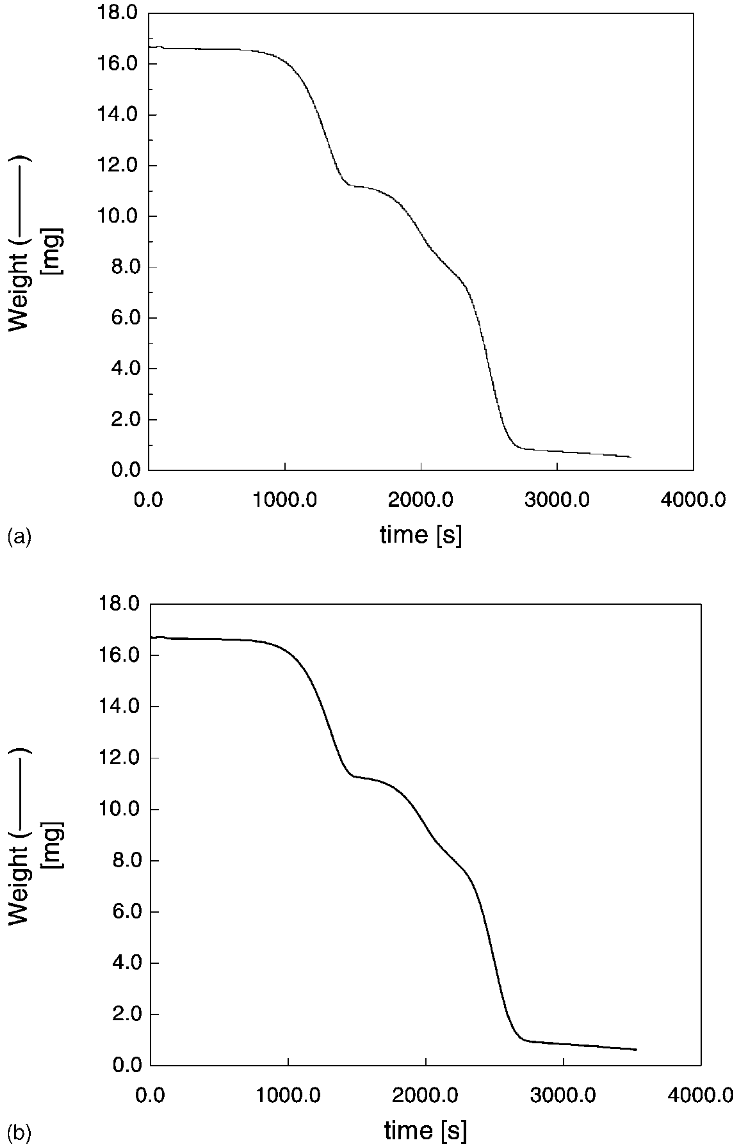
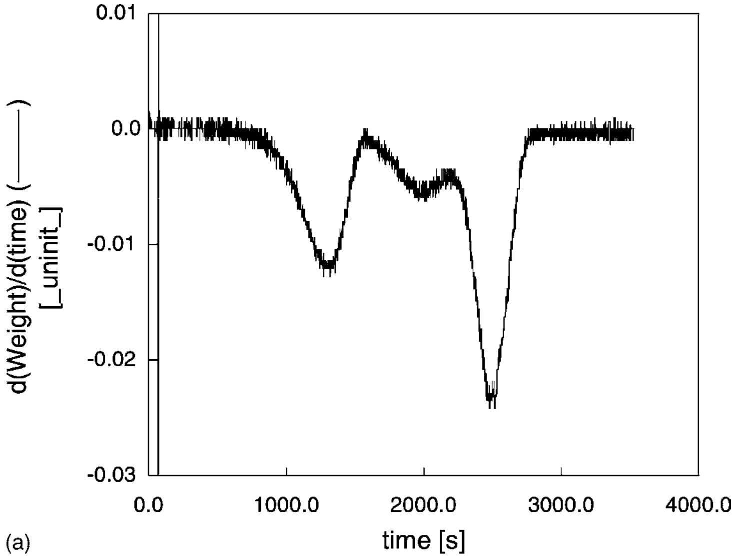


Fig. 3. The TGA raw data obtained from an uncured epoxy resin (a). Local cubic estimation of the TGA trace (b).

than those obtained when using other classical models.

Finally, we considered the data points obtained from the analysis by TGA of an epoxy resin, which thermal behavior was previously described [3]. In this experiment a curing reaction and hardener volatilization occur previously to polymer degradation, giving a complex TGA trace where three weight loss processes, two of them clearly overlapped, are evident. The sample was composed of equal parts by weight of Araldite F and the HY905 hardener. Araldite F is an epoxy resin based on the diglycidil ether of bisphenol A. The hardener is based in a mixture of hexahydrophtalic anhydride, tetrahydrophtalic anhydride, and pthalic anhydride. The both products were supplied by Ciba. A sample of 16.9 mg was placed in open aluminum crucibles. A heating ramp of 10 K/min was applied from 20 to 600 °C and a purge of argon at 50 ml/min was



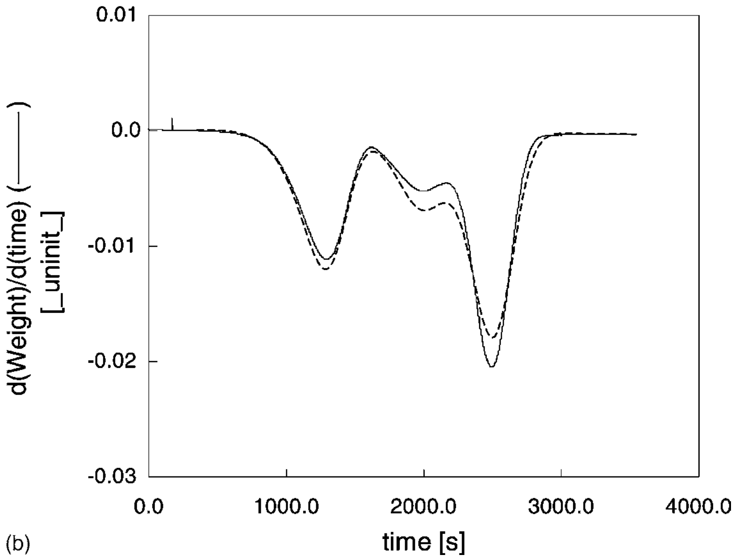


Fig. 4. The TGA first derivative obtained from the apparently smooth raw data by RSI Orchestrator (a). The TGA first derivative obtained and smoothed in Orchestrator (solid line) and a local cubic estimation the TGA first derivative (dashed line) (b).

maintained along the experiment. Fig. 3 plots the estimation of the regression function of these data, with an estimated optimal bandwidth of $\hat{h}_{\text{opt}} = 10.76$, $p = 3$. For comparison, the row TGA data were included in the same figure. Fig. 4 shows the estimation of its first derivative ($\upsilon = 1$) where the bandwidth was $\hat{h}'_{\text{opt}} = 110.63$. As expected, the degree of smoothing to estimate the first derivative is larger than for estimating the function itself. The bandwidth 10.76 accounts for an interval $[t - 10.76, t + 10.76]$ of time (in seconds) around the instant t of interest at which the curve is going to the estimated. This interval needs to the longer: $[t - 110.63, t + 110.63]$ for first derivative estimation. Of course, finally t will range in a wide interval $[0, T]$. Two DTG plots obtained by the Orchestrator software were included in the same figure. The unsmoothed

DTG plot obtained by Orchestrator shows high level of noise. Only for comparison, a smoothing of these data was performed in the same software by choosing the moving average routine, resulting in the solid line. In this case the degree of smoothing was selected by the user looking only at the noise reduction.

5. Conclusions

The parametric logistic regression model proved to be suitable for estimation of TGA derivatives in some simple degradation processes of polymers like a single step in a thermal degradation of PVC.

For more complex reactions that cannot be fitted by a logistic regression model, we propose a model based in a nonparametric estimation, where the goodness of the fit depends very much on the bandwidth selection, especially when the derivatives are concerned. Our method differs from the one proposed by Ruppert et al., in the use of logistic regression (instead of a polynomial regression) for estimating the regression function. The proposed model gave a satisfactory fitting. This way of fitting performs the noise suppression and gives always reliable values, different than the obtained by other methods that depend strongly on the user choosing.

Acknowledgements

The authors gratefully acknowledge for the grant BFM2002-00265 (European FEDER support included).

References

[1] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Chapman & Hall, London, 1997.

[2] D. Ruppert, S.J. Sheather, M.P. Wand, An effective bandwidth selector for local least squares regression, J. Am. Stat. Assoc. 90 (1995) 1257–1267.

[3] J.H. Suwardie, R. Artiaga, L. García, in: Proceedings of the 27th Conference of the North American Thermal Analysis Society, Savannah, 1999, pp. 383–388.