Available online at www.sciencedirect.com
SCIENCE DIRECT®
Thermochimica Acta 412 (2004) 107–111
thermochimica acta
www.elsevier.com/locate/tca

# Competition between kinetic models in thermal decomposition: analysis by artificial neural network

R.C.O. Sebastião, J.P. Braga*, M.I. Yoshida

*Departamento de Química-ICEx, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil*

Received 25 March 2003; received in revised form 24 June 2003; accepted 9 September 2003

**Abstract**

A new approach, based on neural networks, was developed to describe thermal decomposition process. In this method the activation function for the neurons in the hidden layer are substituted by kinetic models functions. Within this framework it was possible to measure the individual importance of the models under consideration.

The rhodium (II) acetate system was used as a prototype to test the efficiency of the neural network. Four models, the Prout–Tompkins model and the Avrami–Erofeev model with $m = 2$, 3 and 4, were selected in a preliminary least square analysis. This will provide the present neural network architecture with an important chemical aspect.

The competition between models was possible to be quantified by the weights in the output layer. Although this thermal decomposition process was, in general, dominated by the Prout–Tompkins model, other models were also important to correctly describe the mechanism. The accuracy of the computed values of decomposition fraction is shown to be greater when compared with the models separately. The present method is of general applicability proposing an alternative efficient way to describe solid thermal decomposition data.


*Keywords:* Kinetic models; Artificial neural network; Solid state thermal decomposition

## 1. Introduction

Solid state thermal decomposition can be described by the quantity $\alpha$, defined as the amount of mass lost at time $t$ normalized to the total mass lost. The process is based on the production and growth of reaction nuclei, the conversion starting at separated points on the crystal surface, associated with imperfections of the crystal lattice [1]. These solid state reactions are described by the kinetic equations that can adequately describe the course of the reaction. Some of the models were proposed by Avrami [2], Erofeev [3], Prout and Tompkins [4], in which kinetic equations, describing the time dependence of decomposition fraction, $\alpha(t)$, were obtained.

Kinetic model of random nucleation followed by linear branched chains growth, the $A_n$ model, was used to describe the thermal decomposition of rhodium (II) acetate [5]. Nevertheless, the correlation between experimental and theoretical results was better described within the neural network framework [6]. Two modifications, change in time scale and correction in the residual amount of the substance, were found to be important.

An alternative method to describe the phenomenon, also using artificial neural network and considering the contribution of several kinetic models, will be discussed in this paper. Although neural network was shown to be a powerful technique to describe the thermal decomposition process, chemical information was lost for more than one neuron in the hidden layer [6]. Nevertheless, the usage of several neurons in the hidden layer cause smaller errors in the neural network procedure. Thus, to recover information about the process, a neural network with several neurons in the hidden layer, keeping fixed the weights in the input-layer, was developed.

This new approach will provide information about the contributions of models that best describe the process at a wide range of temperatures. The accuracy of the computed values will be tested against the best individual model that fits the data.

* Corresponding author. Tel.: +55-31-3499-5742; fax: +55-31-3499-5700.
*E-mail address:* jpbraga@oxigenio.qui.ufmg.br (J.P. Braga).

0040-6031/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.tca.2003.09.009

## 2. Multi-layer neural networks

Neural networks, i.e. softwares that simulate the brain operation, can have architecture with connection only between neurons in consecutive layers. In this case they are termed multi-layer feedforward. The learning process occurs by examples, which consist in giving pairs of input, $\mathbf{t} = [\mathbf{t}_1, t_2, \ldots, t_m]^{\mathrm{T}}$, and output, $\mathbf{o} = [o_1, o_2, \ldots, o_n]^{\mathrm{T}}$. For every input, the network generates the correspondent output, by adjusting the weights in an iterative way [7]. The state of the neuron $j$, that receive these input data, is defined by

$$o_j = f\left(\sum_{k=1}^{l} w_{kj} t_k\right) \tag{1}$$

in which $w_{kj}$ simulates the synaptic connection between the neurons $k$ and $j$ and $f$ simulates the nervous impulse. This process is adapted into a multi-layer neural network.

A simpler neural network can be constructed if some information about the weights for the first layer is given. In this case, and for architecture with two layers, the approach becomes linear, if the output is also linear. The situation is best illustrated in the following steps [8]:

1. The experimental data are multiplied by the weights, $\mathbf{w}_1$, to give $\mathbf{w}_1\mathbf{i}$. Defining, $\mathbf{i} = (\,\mathbf{t} \quad \mathbf{1}\,)^{\mathrm{T}}$ and following Fig. 1 one can define $\mathbf{w}_1$ as

$$\mathbf{w}_1 = \begin{pmatrix} w_{21} & w_{20} \\ w_{31} & w_{30} \\ w_{41} & w_{40} \\ w_{51} & w_{50} \end{pmatrix} \tag{2}$$

to obtain $\mathbf{w}_1\mathbf{i}$ as

$$\mathbf{w}_1\mathbf{i} = \begin{pmatrix} w_{21}t + w_{20} \\ w_{31}t + w_{30} \\ w_{41}t + w_{40} \\ w_{51}t + w_{50} \end{pmatrix} \tag{3}$$

2. The second step, a nonlinear one, requires the neuron impulse to be simulated through a function, the activation function. This function has three important characteristics:
   (a) Before the information has been computed, it assumes a fixed value, generally given by $f(x) = 0$.
   (b) After this, it assumes a value near to 1, implying the neuron has been activated.
   (c) Its first derivative has to be greater or equal to zero. This property will guarantee the neural network will reach a minimum during the training process. This step is represented by the computation of $f(\mathbf{w}_1\mathbf{i})$.
3. The importance of each individual neuron in the intermediate layer is given by the weights in the output layer, $\mathbf{w}_2$. As in Fig. 1 this vector will be defined as, $\mathbf{w}_2 = (\,w_{62} \quad w_{63} \quad w_{64} \quad w_{65}\,)$. The output of the neural network is therefore given by $\mathbf{w}_2 f(\mathbf{w}_1\mathbf{i})$. With a linear activation function in the output layer, the relative importance of the intermediate neurons can be established.

The neural network error, for the present approach, is represented by [7]:

$$E = ||\mathbf{w}_2 f(\mathbf{w}_1\mathbf{i}) - y||_2^2 \tag{4}$$

in which $y$ is the experimental data. The weights $\mathbf{w}_1$ and $\mathbf{w}_2$ have to be optimized to give the minimum error. This calculation can be more efficient if some previous knowledge about the problem is available.

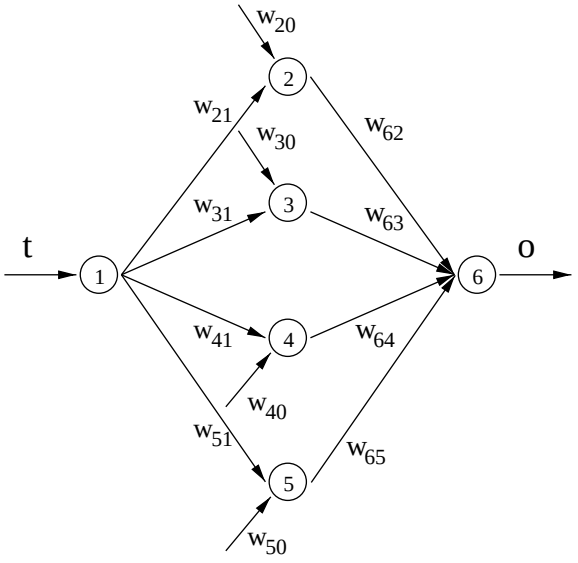


Fig. 1. Neural network architecture.

## 3. Kinetic models and neural network

Theoretical description of thermal decomposition kinetics is based on the formation and growth of reaction nuclei. Therefore, the decomposition rate, $\mathrm{d}\alpha/\mathrm{d}t$, is proportional to the number of active nuclei, that is, to the number of accessible sites ready to decompose, $\alpha$. At the end of the reaction, where no more process will occur, the decomposition rate is proportional to $1 - \alpha$. This will give the right boundary condition since $\mathrm{d}\alpha/\mathrm{d}t$ is zero at the beginning and at the end of the decomposition. Also, if initially there is no active sites, $\alpha(t = 0) = 0$, no processes will occur. In this case mass fraction decomposition will always remain equal to zero. From simple arguments one can establish that $\mathrm{d}\alpha/\mathrm{d}t$ is proportional to $\alpha(1-\alpha)$. With the proportionally constant denoted by $k$, the probability of decomposition will be given by $(1/k)(\mathrm{d}\alpha/\mathrm{d}t)$.

In a more general view, powers of the active and remaining sites are more appropriate to correctly describe the

decomposition kinetics

$$\frac{1}{k}\frac{d\alpha}{dt} = \alpha^{1-m}(1-\alpha)^{1-n} \quad (5)$$

The parameters $m$ and $n$ will determine if the decomposition process is dominated by the acceleratory or retardation stages. It has been showed [9] that all kinetic models can be brought into Eq. (5). It is worthwhile observing this equation is separable in the variables $t$ and $\alpha$, if the integration is to be performed. The integration constant will be denoted by $k_0$.

Integration of Eq. (5) gives for $\alpha$ a function which is similar, mathematically, to a nervous impulse. At the beginning of the decomposition, $\alpha$ is zero, and at the end, it is near to one. This is precisely what happens when the neuron is activated or not. Proceeding with this analogy, each neuron in the hidden layer will have a different activation function, each one corresponding to a kinetic model. The quantity to be activated, for example, at the first neuron, is $w_{21}\mathbf{t} + w_{20}$, whereas the quantity $kt+k_0$ will be "activated" by the kinetic model function. Therefore, an analogy between $w_{21}$ and $k$, also between $w_{20}$ and $k_0$ can further be made. Input-layer weights will be set up to $k$ whereas bias in the hidden layer will be equivalent to $k_0$. Neural network model to describe thermal decomposition, as used here, was motivated by these similarities.

## 4. Results and discussion

Experimental thermal decomposition data, for rhodium (II) acetate system, at the temperatures 198.0, 202.5, 203.5, 205.0, 207.0, 209.0, and 210.5 °C were taken as a reference system to illustrate the efficiency of this neural network method.

A pre-analysis was performed among the available models [9] to obtain the values $k$ and $k_0$. The parameter $k$ represents the rate constant of decomposition process and characterizes the mechanism of reaction. The contribution of different models suggests the process is governed by several mechanisms.

Based on a linear least square procedure, four models were selected: the Prout–Tompkins model [4], $A_n(x) = 1/(1 + e^{-x})$ and the Avrami–Erofeev model [2], $A_m(x) = 1-e^{(-x)^m}$ with $m = 2$, 3 and 4. These kinetic equations satisfy the criterion 2(a)–(c) imposed on the activation function and can, therefore, be used for this objective. The neurons in the hidden layer will then be activated as

$$f(\mathbf{w}_1\mathbf{i}) = \begin{pmatrix} A_n(w_{21}t + w_{20}) \\ A_{m=2}(w_{31}t + w_{30}) \\ A_{m=3}(w_{41}t + w_{40}) \\ A_{m=4}(w_{51}t + w_{50}) \end{pmatrix} \quad (6)$$

The kinetic models will play the role of the activation function.

Table 1
Individual residual errors for the models (number in parenthesis are for power of 10)

| Temperature (°C) | Model | Model error |
|---|---|---|
| 198.0 | $A_4$ | 8.695(−2) |
| | $A_3$ | 5.035(−2) |
| | $A_2$ | 7.840(−2) |
| | $A_n$ | 5.138(−2) |
| 202.5 | $A_4$ | 1.820(−1) |
| | $A_3$ | 2.068(−1) |
| | $A_2$ | 3.421(−1) |
| | $A_n$ | 8.629(−2) |
| 203.5 | $A_4$ | 1.656(−1) |
| | $A_3$ | 6.243(−1) |
| | $A_2$ | 3.787(−1) |
| | $A_n$ | 6.746(−2) |
| 205.0 | $A_4$ | 1.300(−1) |
| | $A_3$ | 1.535(−1) |
| | $A_2$ | 2.500(−1) |
| | $A_n$ | 8.309(−2) |
| 207.0 | $A_4$ | 1.077(−1) |
| | $A_3$ | 8.770(−2) |
| | $A_2$ | 9.895(−2) |
| | $A_n$ | 6.012(−2) |
| 209.0 | $A_4$ | 1.345(−1) |
| | $A_3$ | 1.519(−1) |
| | $A_2$ | 2.203(−1) |
| | $A_n$ | 8.365(−2) |
| 210.5 | $A_4$ | 9.027(−2) |
| | $A_3$ | 1.141(−1) |
| | $A_2$ | 1.743(−1) |
| | $A_n$ | 6.121(−2) |

Since $\mathbf{w}_1$ is fixed the solution of (4) will be [10]

$$\mathbf{w}_2 = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T y \quad (7)$$

with $\mathbf{B} = f(\mathbf{w}_1\mathbf{i})$. The individual importance of each mechanism, represented by the chosen models, will be established by the relative values of $\mathbf{w}_2$.

Together with the property of describing the whole process as a combination of several mechanisms, an additional interpretation of $\mathbf{w}_2$ based on the asymptotic value of $\alpha$ and time scale corrections, is also possible. For large decomposition time the models give $\alpha = 1$ which does not correspond to the real situation. In addition, these models are also corrected by the introduction of another time scale, known as characteristic time scale, $\tau$ [2], defined as $t = p\tau$, where $p$ is the probability of the initial nucleus to be activated. Therefore an analogy between $p$ and $\mathbf{w}_2$ is further possible.

The residual error of each individual model is presented in Table 1. The process is best described by the Prout–Tompkins model, except for the first temperature where the $A_3$ model is also important. In general, one of these models is taken to describe the thermal decomposition process [5] which is clearly an approximation. The present neural network architecture will attempt to provide a better description of the decomposition process.

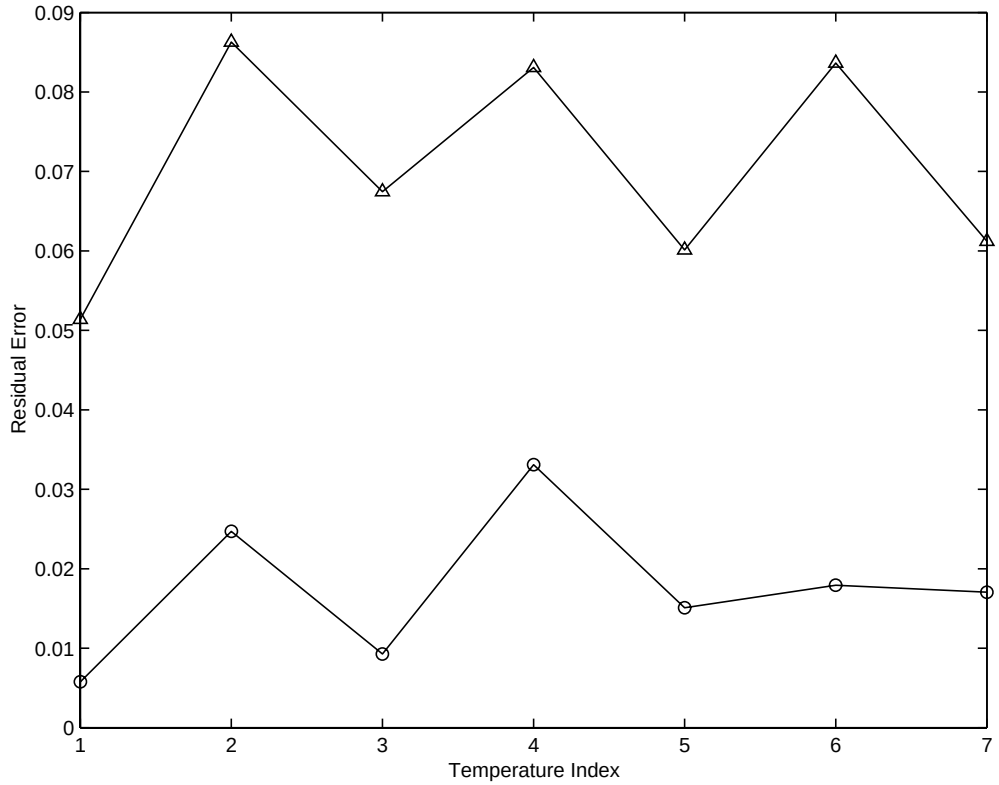


Fig. 2. Residual error for the neural network (○) and $A_n$ model (△). The indexes for the temperatures are 1 for 198 °C, 2 for 202.5 °C, 3 for 203.5 °C, 4 for 205.0 °C, 5 for 207.0 °C, 6 for 209.0 °C, and 7 for 210.5 °C.

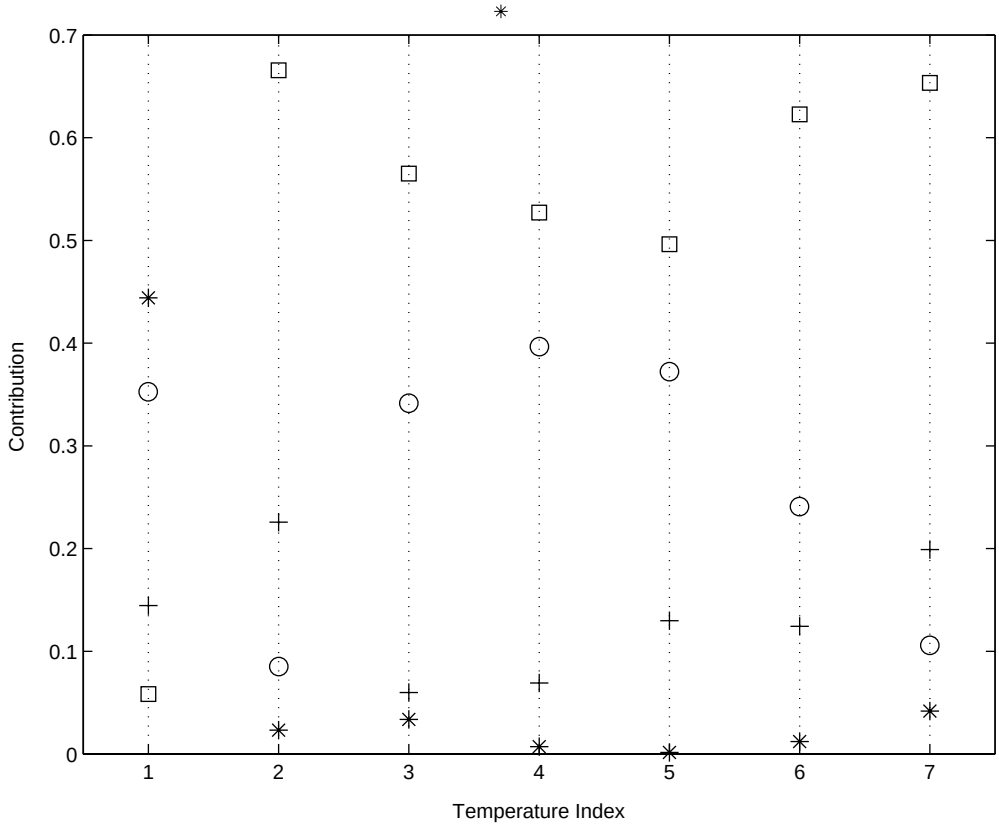


Fig. 3. Kinetic model contribution: $A_n$ (□), $A_2$ (∗ ), $A_3$ (+), $A_4$ (○). Labels for temperatures are same as in Fig. 2.

Fig. 2 presents the residual errors of the individual contribution of the $A_n$ model together with the neural network error (Eq. (4)). The description of the process improves by a factor of 2–7 when it is described by $\mathbf{w}_2 f(\mathbf{w}_1 \mathbf{i})$.

The previous result indicates more than one model is important for the process at a given temperature. Within the present approach this can be quantified by the relative values of the weights in the output layer. The individual contribution is given by

$$c_{6k} = \left| \frac{w_{6k}}{\sum_{j=2}^{5} w_{6j}} \right| \quad (8)$$

for $k = 2, 3, \ldots, 5$ and the results are given in Fig. 3.

The importance of the $A_n$ is evident from this figure, having the same importance as in the individual model analysis. This can be seen by comparing Table 1 with Fig. 3. Nevertheless other models are also important. This is the case for the $A_2$ model at 198 °C, the $A_3$ model for 210.5 °C and the $A_4$ model for the other temperatures.

## 5. Conclusions

The present work discusses an alternative approach to describe thermal decomposition data using artificial neural network. Based on the analogy between the nervous impulse and the time dependence of kinetic models for thermal decomposition, architecture with four models in the hidden layer was constructed. The choice of this number of neurons was made on a preliminary least square analysis among 15 models. Selected models were the Prout–Tompkins model and the Avrami–Erofeev model ($m = 2$, 3 and 4).

Neural networks are powerful techniques to correlate a set of inputs to outputs. A very precise description of the problem can be achieved by controlling the number of layers and the number of neurons in constructing the architecture. Although the neural networks can be a very useful tool for the experimentalist, the chemical contents of the problem are usually lost. That was not the case when a single neuron is used [6] or, as in the present approach, when $\alpha$ is used for the activation function.

Comparing with errors of the models separately, the combination of models in the hidden layer increases the correlation between input and output by a factor of 2–7. Also, within this approach it was possible to quantify the individual importance of the models if the competition between them is considered. The thermal decomposition is better described as a combination of process rather than by an individual mechanism. That was evident from the results.

The method presented in this paper is not restrict to rhodium (II) acetate, and can be applied to other systems, suggesting a powerful routine method to study solid thermal decomposition process.

## Acknowledgements

The authors would like to thank the agencies CNPq and FAPEMIG for financial support.

## References

[1] F.N. Yeremin, The Foundations of Chemical Kinetics, Mir, Moscow, 1979, pp. 383–402.
[2] M. Avrami, J. Chem. Phys. 7 (1939) 1103–1113.
[3] B.V. Erofeev, Comput. Rend. Acad. Sci. USSR 52 (1946) 511.
[4] P.W.M. Jacobs, F.C. Tompkins, Classification and theory of solid reactions, in: W.E. Garner (Ed.), Chemistry of the Solid State, Butterworths/Scientific Publication, London, 1953, pp. 184–211.
[5] M.M. Braga, M.I. Yoshida, R.D. Sinisterra, C.F. Carvalho, Thermochim. Acta 296 (1997) 141–148.
[6] R.C.O. Sebastião, J.P. Braga, M.I. Yoshida, J. Therm. Anal. Cal. 74-3 (2003).
[7] P. De Wilde, Neural Network Models, Springer, London, 1995.
[8] V. Cherkassky, F. Mulier, Learning from Data-concepts, Theory and Methods, Wiley, New York, 1998, pp. 144–149.
[9] W.-L. Ng, Aust. J. Chem. 28 (1975) 1169–1178.
[10] S.J. Leon, Linear Algebra with Applications, 6th ed., Prentice-Hall, New York, 2002.