J Math Chem (2013) 51:1634–1653
DOI 10.1007/s10910-013-0170-7

ORIGINAL PAPER

Combining artificial neural networks and experimental design to prediction of kinetic rate constants

J. L. González-Hernández · M. Mar Canedo · Sonsoles Encinar

Received: 29 January 2013 / Accepted: 19 March 2013 / Published online: 28 March 2013

Abstract A "soft-modelling" computational approach of artificial neural networks (ANNs) combined with experimental design (ED) has been applied successfully in Chemical Kinetics for the prediction of kinetic rate constants. The system studied comprises two consecutive first-order reactions and the kinetic data were computed determining the values of both rate constants. The kinetic curves were distributed according to an ED, and the central star composite experimental design (CSCED) was chosen as the most appropriate. Computational treatments were performed on synthetic data endowed with noise, after which they were applied to the data measured in an experimental reaction between carbonyl cyanide 3-clorophenylhydrazone with 2-mercaptoethanol, computing the experimental kinetic data of absorbance acquired at 3 wavelengths. The combined ANN and ED approach applied in chemical kinetics proved to be robust and of general applicability and has the advantage of being a "soft-modelling" method such that it was not necessary to solve the system of ordinary differential equations to determine the explicit mathematical function between the data and the kinetic rate constants. Additionally, upon using the CSCED experimental design, it was possible to substantially reduce the number of experiments.

Keywords Prediction kinetic constants · Artificial neural networks · Experimental design

J. L. González-Hernández (✉) · M. M. Canedo · S. Encinar
Department of Physical Chemistry, Faculty of Chemistry, University of Salamanca, 37008 Salamanca, Spain
e-mail: jlgh93@usal.es

M. M. Canedo
e-mail: mcanedo@usal.es

S. Encinar
e-mail: sonsoles_e@usal.es

1 Introduction

Computational study of the kinetics of a reaction or a system of chemical reactions is mainly done for two reasons: to determine the kinetic and thermodynamic activation parameters in order to explain a plausible mechanism by which the reaction occurs, and, from an analytical point of view, to achieve a quantitative resolution of homogenous mixtures. The classic treatment used in such studies is the resolution of the system of ordinary differential equations (ODEs) to obtain an explicit mathematical function of the chemical-physical property being monitored with time (usually the absorbance), which includes the kinetic and thermodynamic parameters and initial concentration of the species. This function is different for each kinetic system formed by several species and different chemical reactions, and often involves not a few mathematical problems. Thus, serious problems arise as regards indistinguishability and /or non-unique identifiability, which lead to ambiguities in the solution of ODE system, as occurs in the case of a kinetic model with two consecutive first-order reactions, as studied here, where the function has 2 identical mathematical solutions that lead any curve-fitting method to fail. Even when the kinetic model is slightly more complicated (for example as from three species and four reactions), the ODE system lacks an exact mathematical solution and it is necessary to apply numerical treatments or, sometimes, approximate kinetic methods such as Steady State, Equilibrium State, etc. On other occasions, the function to be minimized has singular points (local minima, saddle points, etc.) that lead even the most robust of mathematical optimization methods to fail.

The computational treatment of the kinetic data for the determination of the kinetic, thermodynamic and analytical parameters is performed using (a) "hard-modelling," approaches [1] where the systems can be exactly described by formulas, equations and the values of parameters, such as mathematical optimization techniques, curve-fitting techniques, Kalman filter algorithms etc. and (b) "soft-modelling" approaches when the exact description is not known or too complex, as multivariate curve resolution (MCR-ALS) [2],where kinetic data relating to absorbance acquired at multi-wavelengths are computed at different times. It is also possible to apply methods that include combinations of parts of both, giving rise to a hybrid approach, as is the case of MCR-PLS [3]. The "soft-" approaches have certain advantages since they act without prior knowledge of the model, such that it is not necessary to solve the ODE system; with this one avoids the mathematical problems of the explicit function mentioned for "hard-" approaches. However, these latter tend to be more robust and are useful and efficient when the kinetic model is known and the function resulting from the resolution of the ODE system is known, is simple, and does not have the afore-mentioned mathematical problems.

In the literature addressing computational kinetic treatment, a series of earlier works [4,5] is of interest where the authors applied different methods of mathematical optimization for the determination of kinetic and thermodynamic activation parameters. These works were the precursors of later ones in which the authors developed different computational programs for application in kinetic, analytical and chemical equilibrium studies KILET [6,7]. To avoid the difficulties found in classic optimization methods, at our laboratory we have designed a new and robust algorithm that has been implemented

in different computational programs (KINAGDC, ANALKIN(AGDC)®, KINMODEL (AGDC),…etc.) developed for the treatment of equilibria [8] and the kinetic study of chemical systems [9–11]. These programs were applied to a broad range of reaction models, successfully determining rate constants, analyte concentrations in multicomponent mixtures (static and dynamic), molar absorption coefficients, etc. Additionally, they allowed us to discriminate between reaction mechanisms, deciding which was the most likely one and the one that best fitted the kinetic data. Later, we applied artificial neural network (ANN) techniques by the ANN application of Matlab [12] for the determination of kinetic parameters. A new approach (ED-ANN) was adapted to the treatment of kinetic concentration data [13], grouped in a particular input matrix ("training" matrix), whose rows jointly contained a sequential ordering of the data and parameters and whose computation was performed with the Simulator Trajan Neural Network and STATISTICA V.6. ANNs treatment of kinetic data has been used in multicomponent kinetic determinations, for analytical purposes [14,15], computing the kinetic data acquired after the reaction of a reagent with the components of the mixture. Moreover, ANN techniques have been used to study the behaviour of chemical reactors in the field of Chemical Engineering [16–18] and the application in Kinetics of an inversion procedure based on recursive neural networks [19].

In the present work we applied an approach combining methods of ANN and experimental design (ED) for the prediction of kinetic rate constants of a system of 2 consecutive first-order reactions. This was a "soft-modelling" approach consisting of application of the trained ANN with the optimum architecture and topology resulting from the "training" process, ale to determine the kinetic parameters without any *a priori* knowledge of the explicit mathematical function that was the solution of the ODE system correlating the data and the kinetic rate constants. The approach was also applied to the measured experimental kinetic data obtained from an experimental reaction taking place in the laboratory.

2 Theoretical aspects

2.1 Chemical kinetic aspects

Let us consider a chemical system formed by n_r chemical elementary (or concerted) reactions where n_s chemical species can be involved. According to IUPAC's norms [20], the system of reaction can be represented:

$$
\begin{aligned}
0 &= \nu_{1,1}B_1 + \nu_{2,1}B_2 + \nu_{3,1}B_3 + \cdots\cdots + \nu_{n_s,1}B_{n_s}\\
0 &= \nu_{1,2}B_1 + \nu_{2,2}B_2 + \nu_{3,2}B_3 + \cdots\cdots + \nu_{n_s,2}B_{n_s}\\
0 &= \nu_{1,3}B_1 + \nu_{2,3}B_2 + \nu_{3,3}B_3 + \cdots\cdots + \nu_{n_s,3}B_{n_s}\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
0 &= \nu_{1,n_r}B_1 + \nu_{2,n_r}B_2 + \nu_{3,n_r}B_3 + \cdots\cdots + \nu_{n_s,n_r}B_{n_s}
\end{aligned}
$$

it can be expressed for the r-th chemical reaction with the generic equation,

$$0 = \sum_{j=1}^{n_s} \nu_{j,r} B_j \tag{1}$$

where

B_j = chemical species involved in the system of reactions
$r = (1, \ldots, n_r)$ number of chemical reactions
$j = (1, \ldots, n_s)$ number of chemical species
k_r = kinetic rate constant of the r-th reaction
$\nu_{j,r}$ = stoichiometric coefficient of the species B_j in the r-th reaction
$\nu_{j,r} < 0$ when B_j plays only the role of reactant in the r-th reaction
$\nu_{j,r} > 0$ when B_j plays only the role of product in the r-th reaction

When the reaction is an elementary or concerted one, the absolute values of the kinetic order ($z_{l,r}$) and stoichiometric coefficient of B_j coincide, that is $|\nu_{l,r}| = |z_{l,r}|$. The rate differential equation of the chemical species B_j in the r-th is given by

$$\frac{d[B_j]}{dt} = k_r \nu_{j,r} \prod_{l=1}^{n_s} [B_l]^{|z_{l,r}|} \tag{2}$$

where B_l are the species playing only the role of reactants in the r-th reaction ($\nu_{l,r} < 0$). Each chemical species can take part in several reactions and the rate differential equations will be the sum extended over those reactions where the reactant B_l appears, obtaining a system of ODEs according to the generic equation,

$$\frac{d[B_j]}{dt} = \sum_{r=1}^{n_r} k_r \nu_{j,r} \prod_{l=1}^{n_s} [B_l]^{|z_{l,r}|} \tag{3}$$

The general solution of the system of rate ODEs give the explicit function of the concentrations of the all species with the time ($[B_j]_{t_i}$).

The chemical system formed by 2 first order consecutive reactions has been studied in the present work and can be represented as

$$\nu_{1,1} B_1 \xrightarrow{k_{12}} \nu_{2,1} B_2$$
$$\nu_{2,2} B_2 \xrightarrow{k_{23}} \nu_{2,3} B_3$$

Considering $\nu_{1,1} = \nu_{2,2} = -1$ and $\nu_{2,1} = \nu_{2,3} = 1$, we have

$$B_1 \xrightarrow{k_{12}} B_2 \xrightarrow{k_{23}} B_3$$

According this, the system of ODE can be expressed using matrix notation [21] as

$$\frac{d}{dt} \begin{vmatrix} [B_1] \\ [B_2] \\ [B_3] \end{vmatrix} = \begin{vmatrix} -k_{12} & 0 & 0 \\ k_{12} & -k_{23} & 0 \\ 0 & k_{23} & 0 \end{vmatrix} \begin{vmatrix} [B_1] \\ [B_2] \\ [B_3] \end{vmatrix} \tag{4}$$

When $[B_1]_0 \neq 0$ and $[B_2]_0 = [B_3]_0 = 0$, the integration of the ODE gives the following expressions of the concentrations for B_1, B_2 and B_3

$$\begin{aligned}
[B_1] &= [B_1]_0 e^{-k_{12}t} \\
[B_2] &= \frac{[B_1]_0 k_{12}}{k_{23} - k_{12}} \left(e^{-k_{12}t} - e^{-k_{23}t} \right) \\
[B_3] &= [B_1]_0 - [B_1] - [B_2] = [B_1]_0 \left(1 - \frac{k_{23}}{k_{23} - k_{12}} e^{-k_{12}t} + \frac{k_{12}}{k_{23} - k_{12}} e^{-k_{23}t} \right)
\end{aligned} \tag{5}$$

According with the Lambert-Beer-Bouguer law, the expression of the absorbance measured, will be:

$$A^{\lambda}_{j,t_i} = \varepsilon^{\lambda}_j \cdot [B_j]_{t_i} \tag{6}$$

where A^{λ}_{j,t_i} is the absorbance of the species B_j at the, time t_i and path length 1 cm and ε^{λ}_j is the molar absorption coefficient of Bj at the wavelength λ. The absorbance of the mixture (A^{λ}_{T,t_i}) can be expressed as:

$$A^{\lambda}_{T,t_i} = \sum_{j=1}^{n_s} A^{\lambda}_{j,t_i} = \sum_{j=1}^{n_s} \varepsilon^{\lambda}_j \cdot [B_j]_{t_i} \tag{7}$$

2.2 Artificial neural networks

Artificial neural networks (ANNs) are parallel interconnected networks of simple computational elements called neurons and are structured in layers that are intended to interact with the objects of the real world in a similar way to the biological nervous systems [22]. Parallel processing is the ability of the brain to simultaneously process incoming stimuli of differing quality. The multilayer neural network uses sets of input data and parameters (called "targets"), distributed in 2 input matrices when Matlab is applied. The elements of the input matrix are the data, where one row contains a single curve and all the curves thus obtained (n_c) are grouped in an "input data" matrix. The "target matrix" is formed by the sets of parameters (n_p). In our case, the input data matrix contained the kinetic data of all curves expressed in A_T, A_j, $[B_j]$ or α_j (molar fraction of the species B_j) and the "target matrix" ($n_c \times n_p$) contained the set of kinetic rate constants (k_{mn}). Formally, a multilayer neural network is an oriented graph in which the nodes represent a set of processing units, called neurons, and the connections represent the information flow channels. Each connection between two neurons has an associated value called "weight" which specifies the strength of the connection between neurons. Positive and negative values determine excitatory and inhibitory connections, respectively. The choice of a specific class of networks for the approximation of a nonlinear map depends on a variety of *factors* dictated by the context and is related to the desired accuracy and the prior information available concerning the input-output pairs.

The first layer of a multilayer neural network contains neurons that receive the input data values from the elements of the input data matrices. This information is transmitted from the i-th neuron of a layer to the j-th neuron of the subsequent one, with a weight w_{ji}. A neuron parameter ("bias") is summed with the weighted inputs of the neurons and passed through the transfer function to generate the output of the neurons.

The layer following the input one is called "hidden". In each neuron of a "hidden" layer the weighed inputs coming from the previous one are summed with each other and added to a "bias". The result is then transformed by means of a suitable mathematical function to obtain an output called "activation of the neuron", which is transferred to the neurons in the next layer after another weighing step. The output parameters values are calculated in the last layer ("output" layer) by means of a suitable transformation function.

The process described is called to as the "training" of the multilayer neural network and constitutes an iterative method where after each iteration ("epochs"), the calculated values of the parameters are grouped in the "output matrix" (b_{ij}^{output}) and they are compared with those of the corresponding curve in the "target matrix" (b_{ij}^{target}) as shown the Scheme 1.

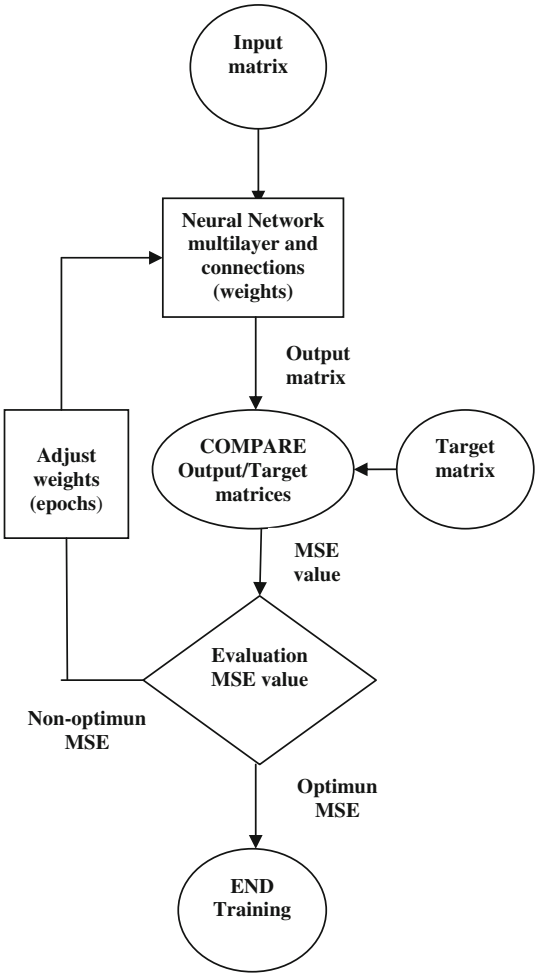
The value of the mean squared error (MSE), expressed in absolute value, is calculated according expression:

$$MSE = \left(\frac{\sum_{i=1}^{n_p} \sum_{j=1}^{n_c} (b_{ij}^{output} - b_{ij}^{target})^2}{n_p \cdot n_c} \right)^{1/2} \tag{8}$$

where n_c is the number of curves and $n_p$is the number of parameters and $n_c \times n_p$ are the dimensions of both matrices (" output matrix" and "target matrix").

During the process of "training", "weights" and "bias" values are modified with suitable mathematical optimization algorithms in order to minimize the calculated values of MSE in each *epoch*. In the present work, the *back-propagation* algorithm was used. The iterative process finishes when the minimum value of MSE is reached, after which the "training" process can be considered to be completed.

It is necessary to know the optimal architecture and topology of the multilayer neural network in order to obtain the best results when ANN is applied to the system under study. This can be performed using the methods described in a recent review [23]. We have used a method of "trial and error" by minimizing the MSE values obtained for the different possible configurations of the same number of "hidden" layer/s. It must to determine the minimum value (optimum) of the MSE for all possible configurations for the "hidden" layer/s chosen. For each "hidden" layer, a graph of MSE values *vs.* the number of neurons *s*hows that initially, for the lower configuration, the value of the MSE decreases rapidly when the number of neurons increases, but after a constant value or a poor improvement is obtained. The optimum number of neurons in that "hidden" layer is given by the point of intersection of the two branches of the graph. Sometimes, a small minimum appears near this intersection point. The architecture of the neural network can be written in abbreviated notation as (n_{inp}, n_{hid}, n_{out}), where



Scheme 1 Flowchart of the iterative method of the optimization of output/target values in the "training" of the neural network

n_{inp} is the number of neurons in the "input" layer, n_{hid} in the "hidden" layer and n_{out} in the "output" layer.

Neural network "training" is completed with the processes of "validation" and "testing". These are 2 control and verification processes of the iterative minimization method between the elements of the "output" and "target" matrices. Among the different curves comprising the "input" matrix, random choice is made of a percentage of the total, established previously (5 %, 10 % ...), which gives rise to a "sub-matrix" of input curves that are subjected to iterative optimization until a minimum MSE value is reached. It is thus possible to verify the validity of the "training" process by ensuring that it is convergent, that it has an appropriate termination, and that there not been any overfitting, since any possible "overtraining" has taken been into account. Validation is completed when in a given number (≥ 6) of consecutive "epochs" the MSE remains constant or shows a slight tendency to increase. The "testing" process is similar, except

that the control of the process is performed by controlling the computation time instead of the number of "epochs".

The process of "prediction" consists of the determination of the unknown parameters from a set of experimental data after application of the optimal and trained neural network. Obviously, the elements of the "target" matrix are unknown for this "prediction" process, and only the "input data" matrix is provided to the neural network. In our case, the elements of the "input data" matrix in the process of "prediction" are experimental kinetic values (A_T, A_j, $[B_j]$, α_j, etc), acquired from a system of reactions developed at the laboratory.

2.3 Experimental design

The experimental design (ED) is a method to select within the working space according to a well-distribution in order to extract the maximum amount of information. The implementation of a suitable ED is crucial to ensure the success of the "training" process of neural networks. In our case, in the ED two *factors* were involved (k_{12} and k_{23}), whose "responses" are the kinetic data of the base of the input curves, which can be of 2 types according to whether they correspond to the contribution of all the species (A_T) or to the individual contribution of a single species (A_j, $[B_j]$ and α_j). In the ED of the neural network it is necessary to consider 2 variables: the limit values of both *factors* that configure the "experimental domain" and their relative values (k_{12}/k_{23}). Both variables must ensure that the binary combinations of the kinetic constants will generate a set of kinetic curves that will have sufficient information to ensure an optimal "training" process of the neural network. Within the interval of time considered, 2 requisites must be reached: (a) the kinetics must achieve a conversion of at least of 80 %($\xi'_{end} = 0.8\xi'_{max}$) and (b) the values of the kinetic data must range within at least 40–50 % so that the neural network will be able to discriminate clearly between the different input curves. Additionally, the number of "levels" of the*factors* of the ED must be suitable if one is to avoid useless computational work and avoid large differences in the spacing of the values of the "responses". Accordingly, to optimize the ANN " training" process the kinetic curves of the "input" matrix must have efficient kinetic information and must be correctly distributed according to the choice of a suitable experimental design and an appropriate "experimental domain". Bearing in mind the characteristics of the kinetic system under study, we have chosen as the most suitable one, the ED corresponding to the *central star composite experimental design* (CSCED) distribution [24] for the combining with the ANNs technique.

3 Computational aspects

The computational work has been performed using Matlab divided in 3 different aspects: (a) the general computational treatment of ANN by means of the application of Matlab ("Neural Networks Toolbox") with the creation of user's interfaces *(GUI)* including the appropriate analysis of Residuals and errors (MSE, SD, etc) (b) design and performing by us of specific computational executable programs (##.m type) in the Matlab environment using "M" language, functions and applications for obtaining

synthetic data, mathematic optimization, fitting functions, etc. and (c) complex 3D graphic representations, endowed of motion and looking for the best perspective in order to obtain an easy visualization of the figure. In the case of the simpler plots other software was used (Origin 8.6 and QTplot)

4 Results and discussion

4.1 Organization of the kinetic data

We considered several types of organizations of the kinetic data in the input matrices to enrich the input kinetic information that would guarantee a more efficient "training" process. The neural network should be endowed with the highest number possible of input options for organizing the data, and this must lead to concordant results. 3-D plotting (A_T/number of curves/time) of the curves corresponding to the A_T data can be seen in Fig. 1a. In the curves of the individual kinetic variables (A_j, $[B_j]$ and α_j) we performed a sequential organization; that is, first the data set corresponding to the species B_1, then those of the species B_2, and finally those of the species B_3 (Fig. 1b). An especially applicable ordering for the kinetic system studied here (3 species and 3 coordinate axes) can be seen in a 3D plot (of the "banana" type) upon representing each individual kinetic variable directly on each of the 3 coordinate axes where all points belong to the same plane (Fig. 1c).

4.2 Training of the ANN. optimal architecture

In order to determine the optimum architecture of the neural network we evaluated the influence of the different variables on the "training" process of the ANN, assessing the effect of each of them in each layer.

4.2.1 "Input" layer

To determine the optimum number of curves we analyzed the ideal number of kinetic data of each kinetic curve. We tested 25, 50, 75 and 100 values, 50 data (50 neurons) proving to be optimum, since the improvement in the results obtained with 75 and 100 values did not justify the huge increase in computation time involved. The ideal maximum range of variation of both *factors* was the same for both, with limit values of 3.12×10^{-2} and $1.19 \times 10^{-1}\text{min}^{-1}$. We varied the number of *levels* and *nodes* of both *factors*, analyzing the responses made. We checked the configurations formed by 29, 37, 45 and 53 nodes or curves, being the optimum configuration proved to be that formed by 45 nodes corresponding to 9 levels and 4 *sub*-levels (Fig. 2). This value offered a sufficient number of *responses*, with no large differences in the spacing, and also is a suitable number of curves for the computational work to be carried out. The dimensions of the *input* matrix were 50×45.

Since the kinetic curves were generated synthetically, the percentages of curves in the "validation" and "testing" processes may be high with respect to the "training" curves. We tested several values, percentages of 80/10/10 proving respectively to be

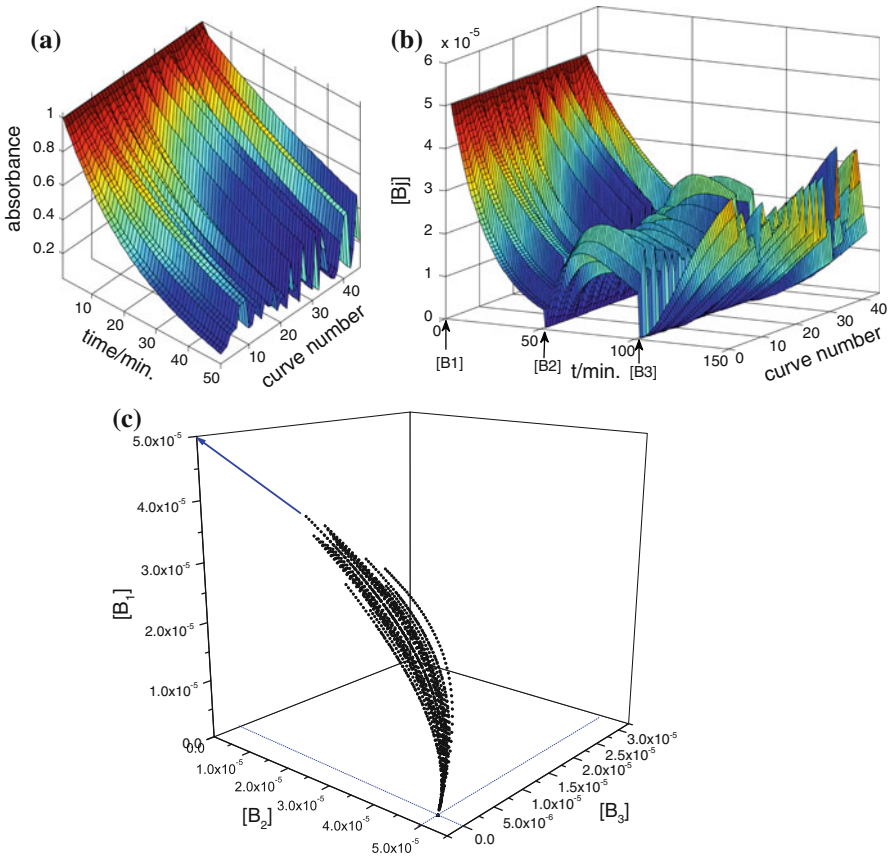


Fig 1 3D plots of the 45 kinetic curves taking part in the curve base of the "input matrix" used for the "training" process of the neural network, **a** each curve contains 50 data of synthetic data of the total absorbance (A_T) endowed with "noise"; **b** 150 data each curve of synthetic data of the total absorbance (A_T) endowed with "noise" with a sequential organization **c** "banana" type plot after representation of the concentrations of the 3 species [B_1], [B_2] and [B_3] on each one of the 3 coordinate axes

optimum for "training", "validation" and "testing", since these percentages ensure a sufficient number of curves for the "training" process. A *random error* (*noise*) must be imposed on the synthetic data, with a magnitude similar to the intrinsic experimental error of the measurement; that is, $\pm 1 \times 10^{-4}$ to the absorbance data and $\pm 1.0 \times 10^{-8}$ to those of concentration ($[B_1]_0 = 5.50 \times 10^{-5}$mol dm^{-3}).

4.2.2 "Hidden" layer/s

Determination of the optimum number of "hidden" layers and that of the neurons of each of them is crucial for the determination of the optimum architecture of the neural network, such that it is necessary to test the highest number of cases possible. The procedure, of the *trial and error* type, consist of initially fixing the number of "hidden" layer/s, carrying out the "training" of the neural network with different configurations,

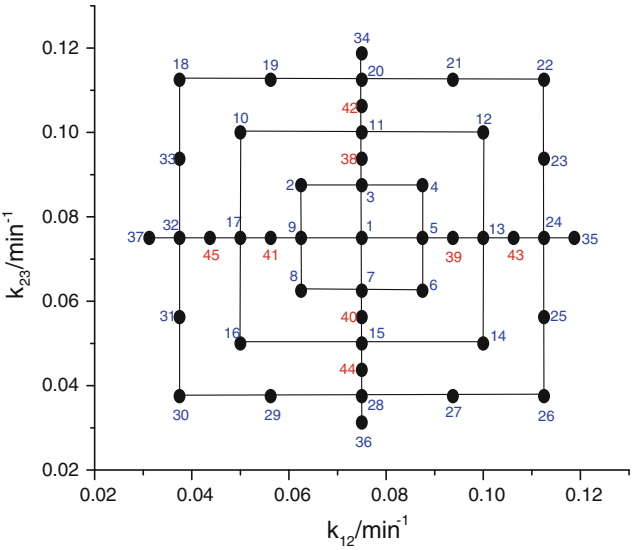


Fig 2 Experimental design (CSCED) constituted by 2 *factors* (k_{12} and k_{23}) and 45 points distributed in 9 levels (points 1–37) and 4 sub-levels: (points 38–45) taking part in the curve base of the "input matrix" used for the "training" process of the neural network

determining the values of MSE, and plotting the values of the MSE values versus the number of neurons for each configuration. As explained above, the observed minimum or the point of intersection of the two branches of the plot, provided (on the abscissa) the optimal number of neurons in that "hidden" layer.

For the A_T input data and a structure of a single "hidden" layer, we obtained an optimal configuration of 5 nodes (Fig. 3a), and 6/10 as the optimal configurations in the case of 2 "hidden" layers (Table 1), obtaining with this latter multilayer structure better results than that of a single hidden layer. In the case of the individual input data (A_j, $[B_j]$ and α_j) and a single "hidden" layer, a configuration of 8 nodes was obtained as the optimum value (Fig. 3b).

4.2.3 "Output" layer

The number of neurons of the "output" layer corresponds to the number of parameters to be predicted; in our case 2. Since Matlab offers the possibility of the application of several optimization algorithms, we tested all of them, choosing as the best the one that provided the best results. We then evaluated the results obtained upon applying the Steepest Descent, Fletcher-Reeves, Scales Conjugate Gradients, Quasi-Newton, BFGS Algorithm and Levenverg-Marquardt (LM), etc. algorithms, the latter (LM) being the one finally chosen since its afforded acceptable MSE values within a reasonable computing time and a suitable number of "epochs".

According with the above we performed the ANN "training" process in the computation of synthetic data endowed with *noise* with the neural network on the basis of the optimal architecture that we obtained, with the following structure: the *input* layer

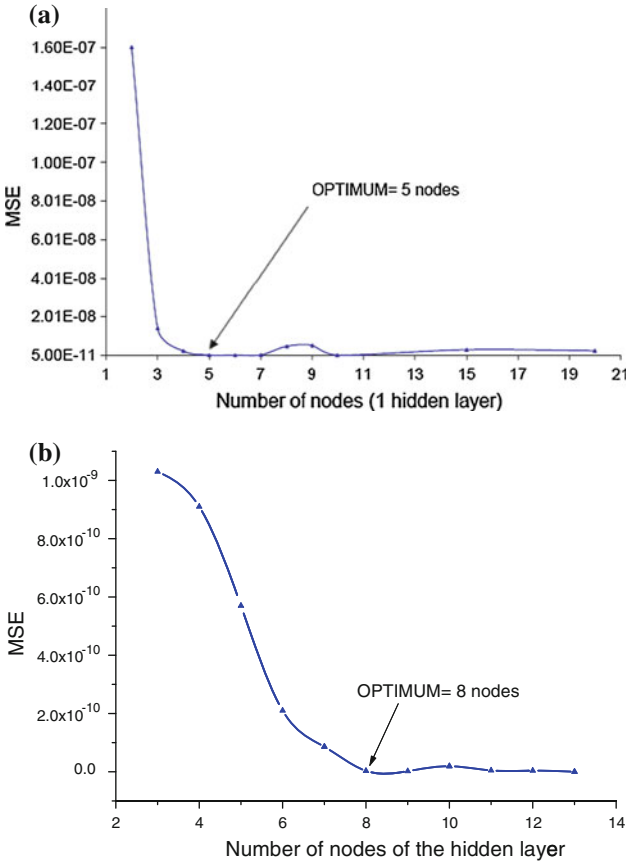


Fig 3 Optimization of the number of neurons (nodes) of the single "hidden" layer, **a** from the computation of individual input data (A_j, $[B_j]$ and α_j) and **b** when A_T input data are computed

and the "output" layer had 50 and 2 neurons, respectively, while the composition of the hidden layer depended on the type of kinetic data of the base of the curves. In the case of individual data (A_j, $[B_j]$ and α_j), the optimum architecture was (50, 1, 2) with a single "hidden" layer, whose suitable configuration had 5 *nodes*.For the A_T data, the optimal architecture was (50, 2, 2) with 2 hidden layers and the configuration of each layer being of 6/10 according with the values of Table 1.

The process of "training" the optimum neural network was performed with the following characteristics; (a) the kinetic data of the "input" matrix formed a base of curves composed of 45 curves, in agreement with a CSCED chosen, with 50 data for each curve and a standard matrix organization in the case of the A_T data and a sequential one for the A_j, $[B_j]$ and α_j data. (b) the *noise* value imposed on the synthetic data $\pm 1 \times 10^{-4}$ for the absorbance data and $\pm 1.0 \times 10^{-8}$ for the concentration data; (c) in all the curves we considered kinetic data with a maximum conversion of no less than 82 % ($\xi'_{end} = 0.82\xi'_{max}$) and the differences between their limit values were always higher than a value of around 50 %; (d) the values of the percentages of

Table 1 Values of MSE for the "training" process of a neural network with 2 "hidden" layers and several configurations

N1	*N*2	MSE
3	8	3.55×10^{-12}
4	10	4.07×10^{-13}
5	7	4.08×10^{-13}
5	10	6.05×10^{-12}
6	7	6.8×0^{-12}
6	9	1.82×10^{-12}
6	10	1.43×10^{-13}
7	6	4.43×10^{-12}
8	6	1.42×10^{-12}
8	7	1.05×10^{-12}
9	7	7.55×10^{-12}
10	5	8.61×10^{-12}
10	10	2.74×10^{-12}

The optimal configuration proved to be 6/10

Table 2 Results of the "training" process of a neural network (50, 1, 2) with a single hidden layer and a configuration of 5 nodes in which the A_j data were computed

BASE: 45 curves, 50 input data
%TRAINING/VALIDATION/TESTING: 80/10/10
ALGORITHM: Levenverg-Marquard
MSE TRAINING: 4.39595×10^{-8}
MSE VALIDATION: 2.45985×10^{-8}
MSE TESTING: 2.36959×10^{-8}
EPOCHS: 100
TIME/s: 0:00:04
GRADIENT: 3.7×0^{-5}
μ: 10^{-7}

" training", "validation" and "testing" chosen were in all cases 80/10/10, and (e) the suitable optimization algorithm used was that of Levenverg-Marquardt (LM).

Table 2 shows the results of the "training" process of a neural network (50, 1, 2) with a single "hidden" layer and a configuration of 5 nodes in which the A_j data were computed.

The profiles of the output/target regression lines are shown in Fig. 4, corresponding to the processes of (a) " training", (b) "validation" (c) "testing" and (d) jointly all processes simultaneously. The results obtained point to satisfactory MSE values (Table 2) and orders of magnitude similar to those of the *noise* errors of the data. The residuals distribution shows acceptable maximum fluctuations and the regression lines have equations with slopes close to unity, significantly null ordinates at the origin, and very good values (0.99998) of the Regression coefficients (R).

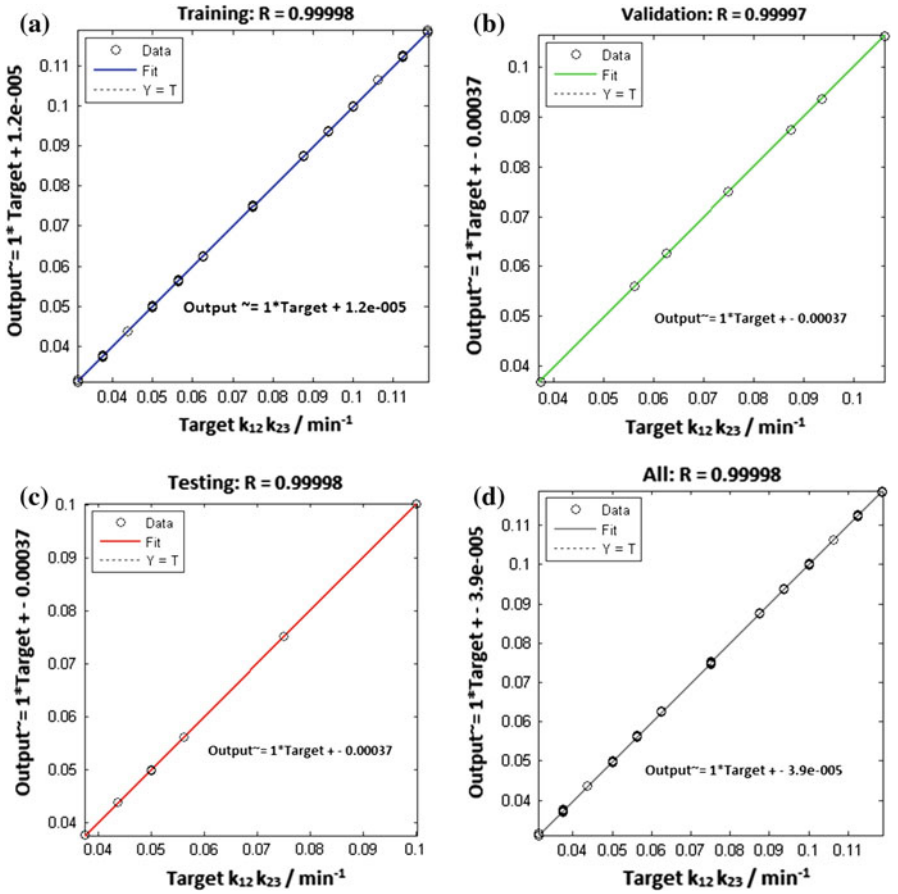


Fig 4 Plots of the representation of the Regression lines outputs/targets of the sets of curves, for the **a** "training" process **b** "validation" process **c** "testing" process and **d** jointly representation for the "training", "validation", and "testing" processes

4.3 Results of the prediction of the ANN

4.3.1 Synthetic kinetic data

In the study of the process of "prediction" of k_{12} and k_{23} from the synthetic data we quantitatively evaluated previously the influence of important variables such as the value of the *noise*, the number of curves to be predicted and the minimum number of species necessary for the "prediction". This is possible with synthetic data, since the values of the parameters are known whereas in the case of experimental data the values of the parameters to be predicted are obviously unknown. The influences studied were as follows:

(a) The *noise* value: we imposed *a priori* different *noise* values on the synthetic data and determined the correlation and the tendency to variation of the error (SD) of

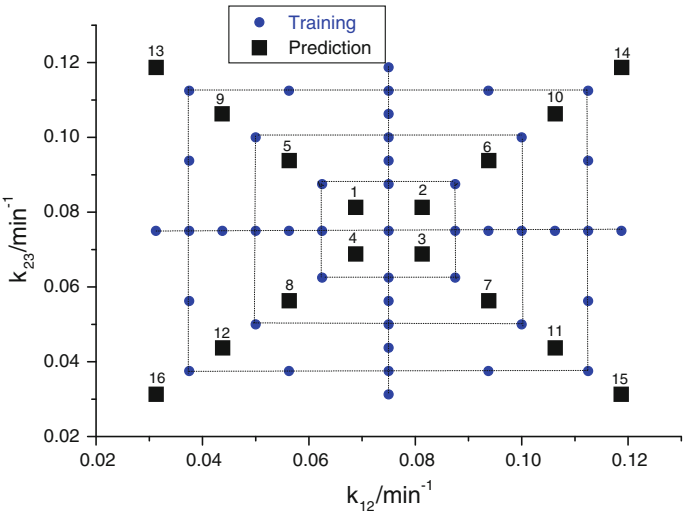


Fig 5 Graphic representation of the Experimental design (CSCED), where the 16 "prediction" points (1–16) are plotted jointly with the 45 points used for the process of "training" of the neural network

k_{12} and k_{23} for the different *noise* error values. It is not necessary to have a rigorous knowledge of the equation of the SD vs. *noise* regression lines; it suffices to know the individual SD values for the working *noise* and of the limits of the "experimental domain" of the *factors* in the maximum range of variation of the ED. These values were:

$$k_{12} = (3.1310^{-2} \pm 3_6 10^{-4})\,\mathrm{min}^{-1} \quad \text{and} \quad k_{23} = (1.18_8 10^{-2} \pm 2_5 10^{-4})\,\mathrm{min}^{-1}$$
$$k_{12} = (1.18_8 10^{-2} \pm 3_6 10^{-4})\,\mathrm{min}^{-1} \quad \text{and} \quad k_{23} = (1.18_8 10^{-2} \pm 2_5 10^{-4})\,\mathrm{min}^{-1}$$

(b) Number of input curves: we determined the optimum number of curves for the "prediction" process, generating synthetic data of A_T, A_j, $[B_j]$ and α_j from points (k_{12}, k_{23}) of the ED different (not coincident) from those used in the "training" process corresponding to points 1–16 of Fig. 5. The plot of the SD values of both rate constants determined in the "prediction" processes on varying the number of curves chosen is shown in Fig. 6. It may be seen that in general upon increasing the number of curves moderately, the value of SD decreases for both rate constants down to a certain threshold value, after which the SD values show a tendency to decrease slightly or remain constant. This value is the optimal one and proved to be in the 8–12 curve range when starting from a base of 45 input curves.

(c) Minimum number of species monitored: we analyzed the possibility of the "prediction" of the rate constants when only 2 species or only a single one contributed to the kinetic data and compared the results with those obtained when the contribution was from the three species. This is a common situation, both when in the spectrum not all the species absorb and when the absorption of one of them predominates over those of the others and the relative values of the molar absorption coefficients are very different. On computing the kinetic data provided by only 2 species, in general

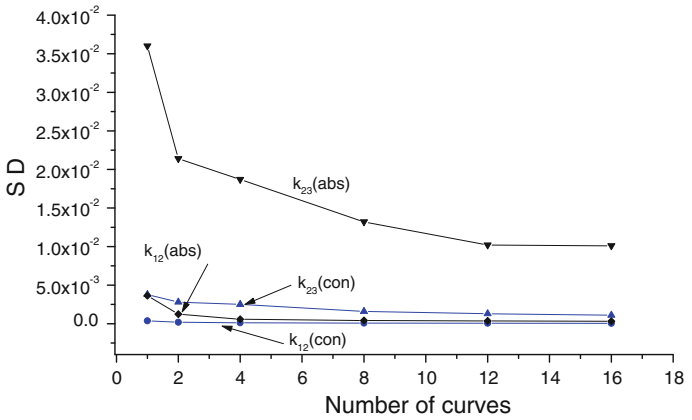


Fig 6 Plot of the SD values obtained in the "prediction" processes of both rate constants (k_{12} and k_{23}) versus the number of curves when both absorbance and concentration data are computed

results similar to those obtained by means of ANN computation of the data in which the 3 species contribute were obtained. This situation was also found in the case of the computation of kinetic data from a single chemical species if this corresponds to the intermediate species (B_2) since it is the one involved in both consecutive reactions.

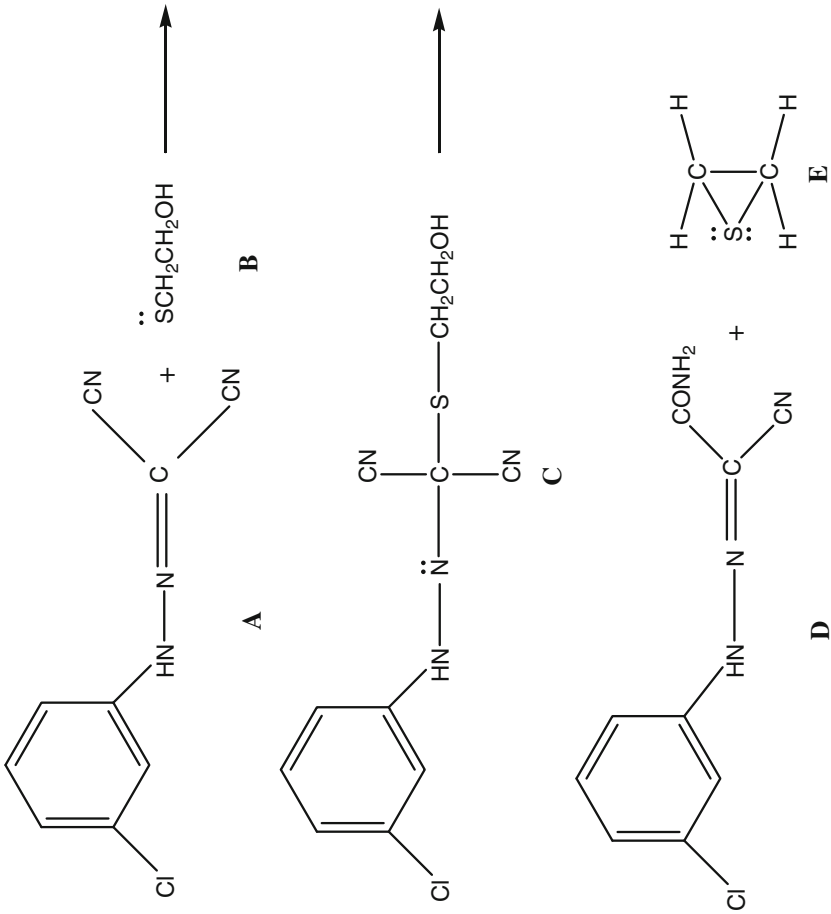
4.3.2 Experimental kinetic data

We performed the ANN treatment to the measured data from an experimental kinetic of a chemical reaction taking place in the laboratory. This was the reaction between *carbonyl cyanide 3-chlorophenylhydrazone* (3-Cl)-PHPD (**A**) *with* 2-mercaptoethanol (**B**). The kinetics of this reaction has been widely studied in the literature [25–27], where the various authors used different classic kinetics methods with concordant results. The reaction occurs according to a kinetic system that involves two irreversible consecutive reactions (Scheme 2). In the first step, an intermediate adduct (**C**) is formed, which is then hydrolyzed in an apparently intramolecular reaction to give the product 3-chlorophenylhydrazonocyanoacetamide, (3-Cl)-PHCA) (**D**) and the subproduct ethylene sulphide (**E**).

Since the reaction is carried out in an excess of the reagent (**B**), pseudo-first-order can be assumed and it is possible to express this schematically according to IUPAC norms [20]:

$$B_1 \xrightarrow{k_{12}} B_2 \xrightarrow{k_{23}} B_3$$

where species (**A**) is represented by B_1, (**C**) by B_2 and (**D**) by B_3. Chau *et al.* [26] performed a classic kinetic study based on data from the total absorbance (A_T) of the reacting mixture, monitored at 3 wavelengths (350, 375 and 400 nm), (Fig. 7) when the reaction conditions are, temperature of 20.0 °C and pH = 4.3. We have used these kinetic data for the computational treatment with the approach combining ANNs and experimental design.



Scheme 2 Mechanism of the reaction between carbonyl cyanide 3-chlorophenylhydrazone (3-Cl)-PHPD with 2-mercaptoethanol

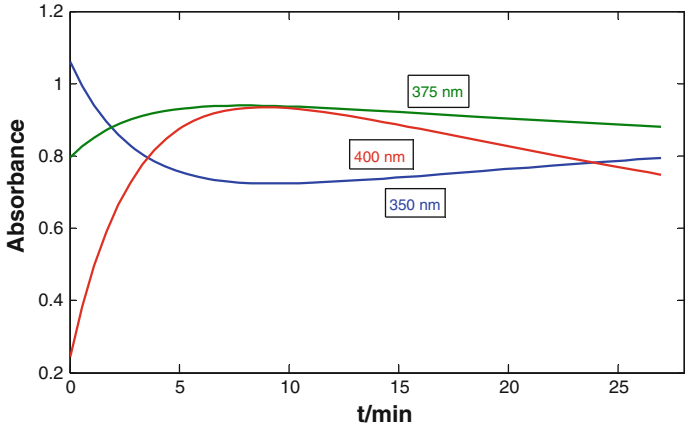


Fig 7 Profiles of the values of A_T with time acquired to 3 wavelengths (350, 375 and 400 nm) for the reaction between (3-Cl)-PHPD with 2-mercaptoethanol

The procedure of the ANN treatment carried out by us has the same steps as in the case of the synthetic data with *noise*; that is, the choice of a suitable ED, "training" of the neural network, evaluating the effect of the variables that provide the ideal conditions of the process and the optimal architecture of the neural network, and, finally, the "prediction" of the unknown values of both kinetic constants (k_{12} and k_{23}) whose values we aimed to determine. The characteristics that made the "training" process optimal were logically almost identical to those used in the case of the synthetic data. Thus, the most suitable "experimental design" was the same; that is, it was of the *CSCED* type. Bearing in mind the values of the initial concentration and those of the molar absorption coefficients of each species at each wavelength, we distributed the 45 curves of the base in 9 levels and 4 sub-levels and an "experimental domain" in the 0.35–0.02 range for both *factors*. As in the case of the synthetic data, each of the 45 curves forming the input base had 50 data of total absorbance (A^{λ}_{T,t_i}), with a significant range of fluctuation ($\approx 50\,\%$) and $\xi_{max} = 0.85\xi$, in all cases using the 80/10/10 curve percentage ratio for "training", validation" and "testing". To determine the optimal architecture of the neural network, must be subjected to a new treatment. To determine the optimal structure and configuration of the hidden layer/s, we first considered an ANN with a single hidden layer (50, 1, 2), assaying a broad range of cases of configurations with different numbers of neurons. In all cases, the MSE values concerning "validation" and "testing" proved to be high (around $10^{-3} - 10^{-4}$) and hence unacceptable. Neither were acceptable the results of the regression lines of the output/target for each configuration nor the "performance" profiles of the curves of the three blocks. Accordingly, it was necessary to rule out the neural network formed by a single "hidden" layer and we tested a neural network formed by 2 "hidden" layers (50, 2, 2) and different configurations for both. The results obtained for the configurations (N1/N2) that provided the minimum validation values at the three working wavelengths (6/10) can be seen in Table 3.

Table 3 Results of neural network formed by 2 hidden layers (50, 2, 2) and different configurations when experimental data of the reaction between (3-Cl)-PHPD with 2-mercaptoethanol acquired at 3 wavelengths, have been computed

N1/N2	λ/nm		
	350	375	400
3/8	6.37E−7	2.40E−8	1.20E−7
5/7	2.58E−10	4.80E−9	1.85E−7
4/10	8.00E−8	1.72E−9	1.24E−7
5/10	2.63E−9	3.41E−9	2.25E−8
6/10	4.31E−9	1.42E−11	5.23E−9

In light of these results, it may be concluded that the optimal architecture of ANNs for the treatment of experimental kinetic data from the reaction under study has two different N1/N2 *configurations*, depending on the working wavelength. Thus, when the absorbance data were measured at $\lambda = 350$nm the neural network was (50, 2, 2), 5/7 being the optimal configuration of the "*hidden"* layers, and when the data were acquired at 375 and 400 nm, the neural network was (50, 2, 2), with an optimal configuration of 6/10. Thus, we were able to predict the unknown values of the rate constants (k_{12} and k_{23}) by applying these two types of architecture and topology of the neural network depending on the value of λ. In all the "predictions", we carried out a study of *Residual Analysis* and *Goodness of Fit*, determining the standard deviation (SD) of the fit with the expression:

$$SD = \left(\frac{\sum_{i=1}^{n_d} \sum_{j=1}^{n_c} \left(A_{ij}^{exp} - A_{ij}^{calc} \right)^2}{n_d . n_c} \right)^{1/2} \tag{9}$$

where A_{ij}^{exp} are the absorbance values of the input matrix and A_{ij}^{calc} are the absorbance values recalculated with the values of the kinetic constants obtained in the "prediction" process; n_d is the total number of data on the absorbance of the input matrix, and n_c is the number of curves.

Table 4 shows the results obtained in the "prediction" process; that is, the values of the k_{12} and k_{23} rate constants together with the SD errors obtained from the Residual analysis after combined treatment of ANNs and Experimental Design of the data on the reaction under study acquired at each of the 3 wavelengths. It may be seen that the results obtained with the "prediction" for the values of the rate constants obtained to the 3 wavelengths point to that coincide satisfactorily. This is proof in favor of the suitability and applicability of the ANN method for the kinetic treatment and consequent determination of the kinetic parameters. The mean values calculated from the individual values at the three wavelengths, for both kinetic constants at 20.0 °C and pH = 4.3, are those corresponding to the kinetics of the reaction studied. These mean values ($< k_{12} >= 0.3381\,\text{min}^{-1}$ and $< k_{23} >= 0.0195\,\text{min}^{-1}$) coincide with

Table 4 Values of the k_{12} and k_{23} rate constants together with the SD errors obtained after the combined treatment of ANNs and experimental design in the "prediction" process of the experimental data on the reaction between (3-Cl)-PHPD with 2-mercaptoethanol acquired at 3 wavelengths

λ/nm	k_{12}/min^{-1}	k_{23}/min^{-1}	SD
350	0.3332	0.020_4	9.26E−4
375	0.3359	0.020_2	7.70E−3
400	0.3454	0.017_9	6.10E−3
450	0.3672	0.031_8	5.92E−4

those obtained by other authors [25–27] who have independently studied the kinetics of this chemical reaction by means of classical kinetic methodologies.

References

1. G. Puxty, Y.M. Neuhold, M. Jecklin, M. Ehly, P. Gemperline, A. Nordon, D. Littlejohn, K. Basford, M. De Cecco, K. Hungerbühler, Chem. Eng. Sci. **63**, 4800–4809 (2008)
2. A. De Juan, E. Casassas, R. Tauler, *Enciclopedia of Analytical Chemistry: Instrumentation and Applications. 'Soft-modelling of analytical data'* (Wiley, New York, 2000)
3. A. De Juan, M. Maeder, M. Martínez, R. Tauler, Chemom. Intell. Lab. Syst. **54**, 123–141 (2000)
4. J. Casado, J.L. González, M.N. Moreno, React. Kinet. Catal. Lett. **33**, 357–362 (1987)
5. J. Casado, J.L. González, M.N. Moreno, G. Sánchez, React. Kinet. Catal. Lett. **36**, 337–344 (1988)
6. J. Havel, J.L. González, P. Palacio, Script. Chem. **22**, 27–33 (1992)
7. J. Havel, F. Cuesta, J.L. González, M.M. Canedo, J. Anal. Chem. **51**, 110–115 (1996)
8. J.L. González-Hernández, M.M. Canedo, A. Domínguez-Gil, J.M. Lanao, J. Pharm. Sci. **81**, 592–596 (1992)
9. M.M. Canedo, J.L. González-Hernández, Chemom. Intell. Lab. Syst. **66**, 63–78 (2003)
10. J.L. González, M.M. Canedo, C. Grande, Int. J. Chem. Kinet. **38**, 38–47 (2006)
11. M.M. Canedo, J.L. González-Hernández, J. Math. Chem. **49**, 163–184 (2011)
12. MathWorks MatLab R2012a, Vs7.14.0.739 (2012)
13. F. Amato, J.L. González-Hernández, J. Havel, Talanta. **93**, 72–78 (2012)
14. S. Ventura, M. Silva, D. Pérez-Bendito, C. Hervás, Anal. Chem. **67**, 1521–1525 (1995)
15. C. Hervás, S. Ventura, M. Silva, D. Pérez-Bendito, J. Chem. Inf. Comput. Sci. **38**, 1119–1124 (1998)
16. S.J. Ardakani, F. Gobal, React. Kinet. Catal. Lett. **85**, 347–382 (2005)
17. P. Valeh-e-Sheyda, F. Yaripour, G. Moradi, M. Saber, Ind. Eng. Chem. Res. **49**, 4620–4626 (2010)
18. B. Kovacs, J. Tóth, Int. J. Appl. Math. Comput. Sci. **4**, 7–11 (2007)
19. N.H.T. Lemes, E. Borges, J.P. Braga, Chemom. Intell. Lab. Syst. **96**, 84–87 (2009)
20. K.J. Laidler, Pure Appl. Chem. **68**, 149–192 (1996)
21. M.N. Berberan-Santos, J.M.G. Martinho, J. Chem. Ed. **67**, 375–379 (1990)
22. Kohonen T, *An Introduction to Neural Computing*, ed. Neural Networks (1988)
23. S. Curteanu, H. Cartwright, J. Chemom. **25**, 527–549 (2011)
24. G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyeres-Verbeke, *Handbook of Chemometrics and Qualimetrics* (Ed. Elsevier, Amsterdam, 1998)
25. R.H. Bisby, E.W.K. Thomas, J. Chem. Ed. **63**, 990–992 (1986)
26. F.T. Chau, K.W. Mok, Comput. Chem. **16**, 239–242 (1992)
27. S. Bijlsma, Louwerse (Ad) D.J., Windig W., Smilde A.K. J. Chemom. **13**, 311–319 (1999)