



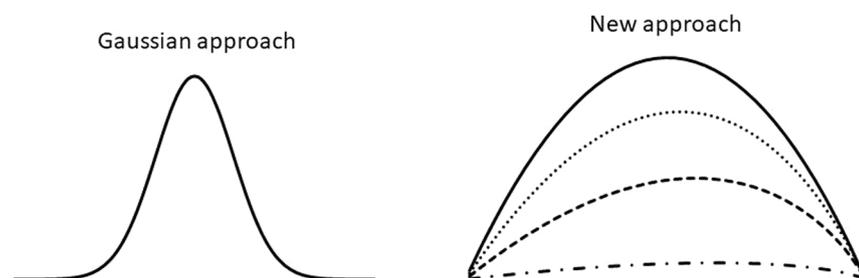
A family of kinetic distributions for interpretation of experimental fluctuations in kinetic problems



Henrique Pacheco, Fabrício Thiengo, Martin Schmal, José Carlos Pinto*

Programa de Engenharia Química/COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária – CP: 68502, Rio de Janeiro 21941-972, RJ, Brazil

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Kinetic error distribution
Experimental error
Chemical kinetics
Confidence interval
Statistical analysis

ABSTRACT

The computation of confidence intervals frequently leads to arguable results due to lack of rigor when experimental errors are analyzed in kinetic experiments. Particularly, the usual Gaussian approach may not be adequate when the variable of interest is the reactant conversion, as this variable is constrained between very hard limits: 0 and 1. For this reason, the present work focuses on the development of analytical and numerical procedures for more accurate description of experimental errors in first-order reaction systems, which can be eventually extended to more complex reaction processes. Based on the proposed analytical and numerical schemes, new statistical distributions (named here as the kinetic distributions) can be derived to allow for more appropriate representation of conversion fluctuations and the respective statistical quantities, including the confidence intervals, which can be used more advantageously for analyses of kinetic data. In particular, it is shown that conversion errors are heteroscedastic, going through a point of maximum when conversion is allowed to increase from 0 to 1, and that confidence intervals are not symmetrical in respect to the averages, as assumed by Gaussian analyses.

1. Introduction

Mathematical models find valuable and widespread use in the field of catalysis. From the most fundamental theoretical aspects to the most complex reaction systems, they are used by researchers to enlighten reaction mechanisms, fit experimental data, and validate proposed hypotheses [1,2]. Besides, models are used in all sorts of kinetic studies, including very different experimental problems, such as chemical vapor

deposition of carbon nanotubes, enzymatic assays, and sewage treatment [3–5].

Kinetic models depend on model parameters that are difficult (not to say impossible) to measure and must be inferred from available experimental data. Definition of model parameters is fundamental during model building because they describe the relative importance of distinct experimental effects on the analyzed process responses. Without proper determination of the model parameters, models become useless.

* Corresponding author.

E-mail address: pinto@peq.coppe.ufrj.br (J.C. Pinto).

<http://dx.doi.org/10.1016/j.cej.2017.09.076>

Received 16 June 2017; Received in revised form 9 September 2017; Accepted 11 September 2017

Available online 15 September 2017

1385-8947/ © 2017 Elsevier B.V. All rights reserved.

Nomenclature

| | |
|----------|--|
| a | Variability of catalytic activity |
| b | First parameter for the variability of concentration measurements |
| c | Second parameter for the variability of concentration measurements |
| C_i | Concentration of species i |
| C_i^e | Concentration of species i at equilibrium. |
| C_{A0} | Initial concentration of species A |
| C_A^m | Measured value of the concentration of species A |
| k_1 | Specific reaction rate for the direction reaction |
| k_2 | Specific reaction rate for the reverse reaction |
| k' | Simplified specific reaction rate |
| k'^m | Measured value of k' |
| N | Number of random numbers to be used in the numerical procedure |
| NE | Number of experiments |

| | |
|-----------------------|---|
| $P_{cum}(x_i)$ | Cumulative probability of variable x at point i |
| S_x | Sample standard deviation of x |
| t | Time of reaction |
| x | Reaction conversion |
| x^m | Measured value of x |
| \bar{x} | Sample mean of x |
| ε_c | Experimental fluctuations of concentration |
| $\varepsilon_{k'}$ | Experimental fluctuations of k' |
| $\varepsilon_C^{(1)}$ | Error measurement at initial concentration |
| $\varepsilon_C^{(2)}$ | Error measurement at equilibrium concentration |
| $\varepsilon_C^{(3)}$ | Error measurement at sample concentration |
| ε_x | Conversion fluctuations |
| $\bar{\varepsilon}_x$ | Measurement bias of x |
| $\mathcal{G}_z(z)$ | Probability distribution of variable z |
| σ_a^2 | Variance of a |
| σ_a | Standard deviation of a |
| μ_a | Mean of a |

When the proposed study requires the evaluation of model parameters, it is common to estimate the unknown parameters through minimization of an objective function (or metrics) that represents the distance between the values predicted by the model and the values obtained experimentally [2]. In order to formulate the objective function, it is usually necessary to postulate several simplifying assumptions, either to reduce the experimental load or simply because it is impossible to know beforehand all features of a particular experimental system. Most times, however, these assumptions are not validated or verified, which may lead to inappropriate conclusions and parameter values.

In order to illustrate this point, Schwaab et al. [6] showed that the confidence region of parameter values may present very peculiar shapes when the nonlinear behavior of model responses and experimental measurements are considered. Such region can be quite distinct from the hyper-ellipsoidal shape commonly assumed for confidence regions, as obtained for linear model responses and Gaussian distribution of measurement fluctuations. Similarly, Schwaab et al. [7–9] showed that improper model parameterization may lead to highly correlated model parameters, which can significantly degrade the final model performance. Finally, it is usually assumed that fluctuation errors are constant throughout the experimental region, even though this cannot be supported by independent error analyses, as discussed by Alberton et al. [10].

Certainly, one of the commonest assumptions is the Gaussian distribution for the experimental fluctuations. This assumption is so popular that many do not understand that, although the Gaussian distribution is a model that may find suitable applications in numerous situations, it may also be inadequate for the interpretation of several other practical problems [2]. For this reason, the proper characterization of experimental fluctuations can be of paramount importance for correct assessment of experimental analyses and interpretation of kinetic models. Despite that, the detailed characterization of experimental fluctuations in kinetic studies is often neglected, due to difficulty to investigate how these unavoidable fluctuations depend upon the reaction conditions, among other reasons. Particularly, the behavior of the experimental fluctuations can be linked to specific characteristics of the experimental system, including measuring techniques and operation procedures, naturally causing this type of investigation to be challenging [2].

As experimental fluctuations are related to uncontrolled random causes, the proper characterization of experimental variability requires replication of experimental trials a sufficiently high number of times [11], discouraging the fundamental investigation of fluctuations and explaining why certain statistical distribution models are needed and

assumed to be valid *a priori* during an experimental investigation. Particularly, the Gaussian distribution is useful because it may be applied to a large array of physical problems, provides an asymptote for problems dominated by infinitely many sources of variability and requires the definition of only two parameters (mean and variance) for its use. Moreover, the Gaussian distribution is mathematically tractable, allowing for a number of important theoretical developments, which include the derivation of t-Student, F-Fisher, and chi-square tests for analyses of means and variances of experimental data samples [2,11].

When the Gaussian distribution is used to define boundaries for conversion and selectivity measurements in kinetic studies, however, anomalous results may be obtained. For instance, confidence intervals may lie outside the [0,1] interval (which makes no physical sense), because the Gaussian distribution is defined in the infinite domain, while conversions and selectivities lie in the much narrower finite interval [0,1]. An obvious conclusion is that conversion and selectivity measurements do not follow the Gaussian distribution, although it may be true that this distribution may provide useful fits for experimental fluctuations in certain experimental systems.

Based on the previous paragraphs, the present work focuses on the development of analytical and numerical procedures for more accurate description of experimental errors in first-order reaction systems. Based on the proposed analytical and numerical schemes, new statistical distributions (named here as the *kinetic distributions of fluctuation measurements*) are derived to allow for more appropriate representation of conversion fluctuations and the respective statistical quantities, including the confidence intervals, which can be used more advantageously for analyses of kinetic data. In order to do that, a first-order reaction is assumed to take place in a model reacting system, as several systems can be represented with good accuracy by first-order reaction models. Besides, it is well known that more complex nonlinear functions can be represented locally by simpler models, especially when experimental fluctuations are not too large [12]. Nevertheless, as shown in the proposed numerical development, this underlying assumption does not constitute a major drawback of the proposed analysis, for the first-order reaction rate assumption can be easily relaxed in more involving numerical analyses in order to represent more complex reaction systems.

Stochastic methods are largely applied in Chemical Engineering, both as an alternative for finding global maxima and as an elegant, efficient way to validate simulations [13–15]. What is proposed in this work is, though, to explore the error distributions themselves via a stochastic approach, something that is seldom seen in the scientific literature. This is of the utmost relevance since errors in the most fundamental variables of the problem must have their statistical

probability density distribution properly described as a starting point for the application of the maximum likelihood approach in parameter estimation problems, for instance. Should this primordial error distribution (widely assumed to be Gaussian) be incorrect, all the remaining estimation process is tainted and the final parameters will carry an extra level of uncertainty. It is also quite important, on the other hand, to keep in sight that the proposed methodology does not aim to build the “true” error distribution of the system. Nevertheless, the fact is that this “true” distribution is not really required to perform the error analysis, but only a fair approximation of it. Such hypothesis is strongly supported by the widespread use of the Gaussian distribution in several problems, despite it not being the correct error distribution of most of them. The core idea of this work is to propose an error distribution that is more appropriate for statistical analyses of kinetic data than other distributions, since it satisfies relevant experimental constraints and is derived from an actual kinetic problem. Even if it does not describe the “true” error distribution, one must acknowledge that such perfect distribution will never be known in a real experimental kinetic problem anyway.

2. Conversion fluctuations in a first-order reacting system

Let a generic reaction mechanism be described by Eq. (1):



where A and B represent the reactant and the product and k_1 and k_2 represent the specific reaction rates for the direct and reverse reactions, respectively. If the reaction is conducted isothermally in a reaction system without significant volume variation, the material balances can be written as described by Eqs. (2) and (3):

$$\frac{dC_A}{dt} = -k_1 C_A + k_2 C_B; \quad C_A(0) = C_{A0} \quad (2)$$

$$\frac{dC_B}{dt} = k_1 C_A - k_2 C_B; \quad C_B(0) = 0, \quad (3)$$

where C_A and C_B are the concentrations of A and B , respectively. At the equilibrium condition it can be stated that:

$$k_1 C_A^e = k_2 C_B^e \Rightarrow k_1 C_A^e = k_2 (C_{A0} - C_A^e) \Rightarrow C_A^e = \frac{k_2 C_{A0}}{k_1 + k_2}, \quad (4)$$

where the superscript e represents the concentration at equilibrium. Normalized conversion (normalized in respect to the species' concentration at equilibrium) may be expressed as:

$$x = \frac{C_{A0} - C_A}{C_{A0} - C_A^e} = \frac{C_B}{C_B^e} \quad (5)$$

so that x is defined within the $[0,1]$ range and describes the conversion of A as a function of the maximum attainable conversion. Introducing this variable into Eq. (2):

$$\frac{dx}{dt} = -(k_1 + k_2)x + k_1 \frac{C_{A0}}{(C_{A0} - C_A^e)}; \quad x(0) = 0, \quad (6)$$

whose solution is given by Eq. (7):

$$x = 1 - \exp[-(k_1 + k_2)t] = 1 - \exp(-k't) \quad (7)$$

Solving for t and isolating C_A :

$$t = -\frac{\ln(1-x)}{(k_1 + k_2)} = -\frac{\ln(1-x)}{k'} \quad (8)$$

$$C_A = C_{A0} - (C_{A0} - C_A^e)[1 - \exp(-k't)] \quad (9)$$

It is important to observe that Eq. (9) involves one process response (C_A), two experimental inputs (C_{A0} and t) and two model parameters (C_A^e and k'). More complex kinetic problems will certainly involve additional sets of response variables, input variables, and model parameters. In this case, an analytical solution for the model responses may

not be available, although mass balances can always provide the desired solutions after implementation of suitable numerical procedures. This is why the reader should not regard Eq. (9) as a fundamental constraint for the proposed analysis, as numerical responses can always provide numerical solutions for more involving kinetic models and reacting systems.

Even when the model is assumed to provide accurate description of the analyzed kinetic problems (in this case, the evolution of conversions with time), model responses and experimental measurements are not equal due to the inevitable occurrence of measurement errors. It is assumed here that the two main sources of experimental fluctuations are the concentration measurements (usually via chromatographic analyses) and the catalytic activity (due to disturbances in preparation conditions, structure of the catalytic bed, particle morphology, among others). These sources of fluctuation can be described according to Eqs. (10) and (11):

$$C_{A0}^m = C_{A0} + \varepsilon_C^{(1)}$$

$$C_A^{e,m} = C_A^e + \varepsilon_C^{(2)} \quad (10)$$

$$C_A^m = C_A + \varepsilon_C^{(3)}$$

$$k'^m = k' + \varepsilon_{k'}, \quad (11)$$

where m represents the disturbed value, ε is a measure of the experimental fluctuation and the superscripts (1), (2) and (3) represent the measurements of error at initial, equilibrium, and sample concentrations, respectively. Consequently, one can write down the equation for the normalized conversion as:

$$x^m = \frac{C_{A0}^m - C_A^m}{C_{A0}^m - C_A^{e,m}} = \frac{(C_{A0} + \varepsilon_C^{(1)} - C_A - \varepsilon_C^{(3)})}{(C_{A0} + \varepsilon_C^{(1)} - C_A^e - \varepsilon_C^{(2)})}, \quad (12)$$

This way, conversion fluctuations (ε_x) can be written in the forms of Eqs. (13) and (14):

$$\varepsilon_x = x^m - x = \frac{(C_{A0} + \varepsilon_C^{(1)} - C_A - \varepsilon_C^{(3)})}{(C_{A0} + \varepsilon_C^{(1)} - C_A^e - \varepsilon_C^{(2)})} - [1 - \exp(-k't)] \quad (13)$$

$$\varepsilon_x = x^m - x = \frac{(C_{A0} + \varepsilon_C^{(1)} - \{C_{A0} - (C_{A0} - C_A^e)[1 - \exp(-k't)\}] - \varepsilon_C^{(3)})}{(C_{A0} + \varepsilon_C^{(1)} - C_A^e - \varepsilon_C^{(2)})} - [1 - \exp(-k't)] \quad (14)$$

Again, it is important to observe that Eq. (14) shows how measurement errors are correlated with and depend on the variable inputs and model parameters. Obviously, analytical expressions may not be available in more involving kinetic problems, but can always be computed numerically for a particular model structure and set of model inputs and parameters.

Using Eq. (8), Eq. (14) can be rearranged as:

$$\varepsilon_x = (1-x) \left\{ 1 - \exp\left(\ln(1-x) \frac{\varepsilon_{k'}}{k'}\right) \left[1 - \frac{(\varepsilon_C^{(1)} - \varepsilon_C^{(2)})}{C_{A0} - C_A^e} \right] \right\} + \frac{\varepsilon_C^{(2)} - \varepsilon_C^{(3)}}{(C_{A0} - C_A^e)} \quad (15)$$

$$\varepsilon_x = (1-x) \{1 - \exp[\ln(1-x)a][1-b]\} + c, \quad (16)$$

where parameter a describes the variability of catalytic activity and parameters b and c describe the variability of the concentration measurements, respectively. Generally, it is more difficult to control all the many operation parameters related to catalyst preparation, catalyst bed morphology and particle shape than to control concentration measurements. As a consequence, it seems reasonable to assume that small variations of catalyst activity control the experimental accuracy, becoming possible to express conversion fluctuations as a function of catalyst activity variability (a):

$$\varepsilon_x = (1-x) \{1 - \exp[\ln(1-x)a]\} \quad (17)$$

When, instead, fluctuations of concentration measurements control

the experimental fluctuations, then:

$$\varepsilon_x = (1-x)b + c \quad (18)$$

Although Eqs. (17) and (18) are very simple, it is important to observe that conversion measurement fluctuations depend in both cases on the conversion levels, as previously observed by Alberton et al. [10], so that conversion variances change along the experimental region (the system is heteroscedastic). Besides, Eq. (17) shows that there is a nonlinear dependence between conversion and catalyst activity variabilities, which should not be disregarded during the analysis of the experimental errors. As pointed out previously, Eqs. (17) and (18) can be relaxed to accommodate different reacting systems and hypotheses, although analyses will have to rely on numerical computations in such cases.

3. A kinetic probability distribution for conversion measurement fluctuations

It is important to observe that, even when experimental fluctuations of the fundamental variables (catalyst activity and concentration measurement) can be assumed to follow the Gaussian distribution, the equations that describe conversion measurement fluctuations indicate that variability of conversion data do not necessarily follow the Gaussian model. This is not surprising, as x is defined in the [0,1] range, while the Gaussian distribution is defined in the infinite domain, as already stated. Therefore, in order to correlate the fluctuations of catalyst activity and concentration measurements to the uncertainties of x , it can be assumed that:

$$\varphi_{\varepsilon_x}(\varepsilon_x) = \varphi_a(a) \frac{da}{d\varepsilon_x} \quad (19)$$

$$\varphi_{\varepsilon_x}(\varepsilon_x) = \int_0^{+\infty} \varphi_c(c) \varphi_b(b(c;\varepsilon_x)) \frac{\partial b}{\partial \varepsilon_x} \Big|_c dc \quad (20)$$

where $\varphi_z(z)$ is the probability distribution of variable z . Eqs. (19) and (20) assume that the probability to find a conversion measurement in a differential interval dx is equal to the probability of finding the catalyst activity in the differential interval da or to the probability of finding concentration measurements in the differential interval dc , depending whether fluctuations of catalyst activity (Eq. (19)) or concentration measurements (Eq. (20)) control the conversion measurement fluctuations. Assuming that the normal distribution can be used to represent uncertainties of the catalyst activity fluctuations, in the form of Eq. (21):

$$\varphi_a(a) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{(a-\mu_a)^2}{2\sigma_a^2}\right), \quad (21)$$

where σ_a^2 is the variance of catalyst activity (or catalytic activity variability) and μ_a is its mean, then it is possible to write Eq. (19) as:

$$\varphi_{\varepsilon_x}(\varepsilon_x) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{(a-\mu_a)^2}{2\sigma_a^2}\right) \frac{1}{[(1-x)-\varepsilon_x] \ln(1-x)} \quad (22)$$

An analogous procedure can be applied to Eq. (20) when fluctuations of concentration measurements control the experimental fluctuations.

At this point, it is important to highlight that the Gaussian distribution presents an intrinsic limitation for representation of catalyst activity fluctuations, as activities cannot be negative. When a becomes smaller than -1 in Eq. (16), the reaction is indeed reversed, as the kinetic constant k' becomes negative (the negative fluctuation of catalyst activity becomes larger than the nominal catalyst activity). For this reason, other reference distributions can be used to describe fluctuations of catalyst activities, such as the uniform distribution. Additional information regarding the influence of the reference distribution on the proposed numerical procedures is presented in Appendix A.

Fig. 1 presents the flowchart that describes the numerical procedure that can be used to build the probability distribution of conversion measurements (the kinetic probability distribution) due to catalytic activity fluctuations and to define the boundaries for conversion fluctuations. Initially, the standard deviation for fluctuations of catalytic activity (σ_a) must be defined, assuming that the average fluctuation value is equal to zero. Other types of variability measurements can be used, although the standard deviation is certainly the most used measure of process variability. Then a sufficiently large number (N) of random numbers must be generated (5000 points were used in this work), according to the appropriate reference distribution (in the present work, unless stated otherwise, the Gaussian distribution is used as reference – respecting the limitation that $a > -1$ –, although this assumption is not fundamental and can be easily relaxed, as discussed in Appendix A). The next step is to compute values of a with the N random numbers using the value for σ_a already defined. Then, a reference conversion fluctuation value must be obtained, either experimentally (as average of sampling values, as described below) or calculated with the parameters of the distribution (k' and C_{A0}). With the established values of a and x , one must then use them to estimate the conversion deviations (ε_x) with help of Eq. (17) (or Eq. (18), if conversion fluctuations are controlled by concentration measurement fluctuations). Finally, the cumulative kinetic probability distribution of ε_x can be obtained, assuming that [2]:

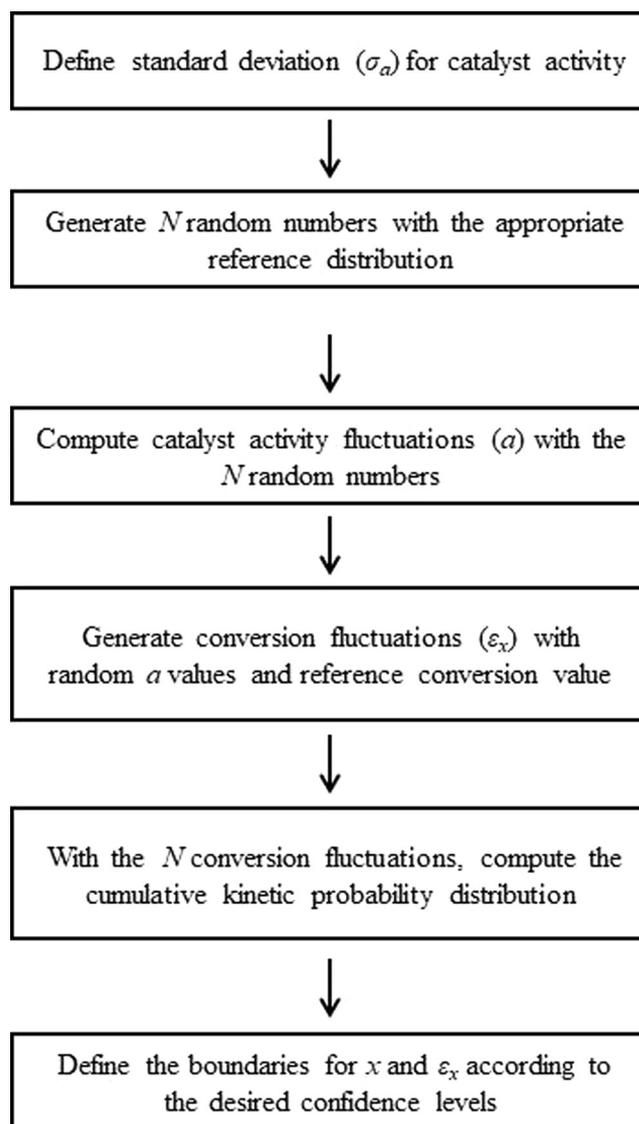


Fig. 1. Procedure for definition of the kinetic probability distribution.

$$P_{cum}(x_i) = \begin{cases} \frac{1}{N+1}, i = 1 \\ P_{cum}(x_{i-1}) + \frac{1}{N+1}, i = 2, 3, \dots, N \end{cases} \quad (23)$$

Eq. (23) can be used to compute the first moments of the kinetic distribution (mean and standard deviation of x and ε_x) and the boundaries of conversion values and conversion fluctuations, using the desired confidence level (unless stated otherwise, the confidence level of 95% is used below).

Fig. 2 illustrates how the catalyst activity affects conversion fluctuations, for distinct conversion values. As one might expect, as the variability of catalytic activity increases, the magnitude of the conversion fluctuations also increases for all reference conversion values. However, it is important to observe that conversion fluctuations can change considerably within the experimental range, going through a point of maximum that depends on the fluctuations of the catalyst activity. Besides, it is also important to notice that the nonlinear function is not symmetric in respect to x , changing faster in the vicinities of $x = 1$. Similar behavior had been observed for experimental conversion variances in first-order reacting systems by Alberton et al. [10]. It is very relevant to highlight that conversion fluctuations are equal to zero at the boundaries $x = 0$ and $x = 1$, as one might expect from physical grounds, since real conversion values cannot surpass these hard experimental constraints. (When conversion fluctuations are controlled by concentration measurement fluctuations, they can be eventually violated by measurement of wrong concentration values).

Fig. 3 presents the cumulative kinetic distribution as a function of ε_x , at the reference conversion value of 0.5 and three different levels of catalyst activity variability. This graph displays some interesting characteristics of the kinetic distribution. First of all, it can be noticed that the cumulative distribution of ε_x is not symmetrically distributed around zero, given the nonlinear nature of the correlation among the many analyzed variables. An interesting consequence of this point is that sample averages of experimental conversion values may be biased and different from the real reference value, which has been completely overlooked in previous publications and disregarded in experimental kinetic studies. Besides, the shape of the cumulative kinetic distribution can be significantly affected by the variability of catalyst activity, being distorted towards higher values of ε_x when σ_a increases.

Table 1 shows the boundaries of ε_x for a reference conversion value of 0.5, illustrating how the asymmetrical behavior varies when σ_a changes. Based on results presented in Table 1, it seems clear that confidence regions provided by the Gaussian distribution may constitute very poor approximations for conversion values obtained experimentally. Besides, Table 1 also shows that proper characterization of experimental fluctuations must somehow take into account the natural variability of catalyst activity in kinetic studies (a more complete list of boundary values, for different conversion levels and catalyst variabilities, is presented in Table S1 as Supporting Data).

Table 2 displays how the first moments of the kinetic probability distribution (mean and sample standard deviation values for conversion) depend on catalyst activity variability (σ_a) for several levels of conversion (a more complete list of means and standard deviations, for different conversion levels and catalyst variabilities, is presented in Table S2 as Supporting Data).

4. Practical use of the kinetic probability distribution

From a practical point of view, the potential user of the kinetic probability distribution may not be willing to perform the proposed computations, but may be convinced at this point that confidence regions might be calculated more accurately using the proposed kinetic distribution, instead of the more usual Gaussian distribution. In this case, it is important to observe that the user usually relies on a sample of NE experiments, which can be used to provide experimental sample means and sample standard deviations for the reaction conversion, calculated respectively as:

$$\bar{x} = \frac{1}{NE} \sum_{i=1}^{NE} x_i \quad (24)$$

$$S_x = \sqrt{\frac{1}{(NE-1)} \sum_{i=1}^{NE} (x_i - \bar{x})^2} \quad (25)$$

Assuming that the standard deviation is free of error (this assumption will be relaxed below), then Table 2 (or Table S2) can be used to provide the values of the catalyst variability standard deviation (σ_a) and of the measurement bias ($\bar{\varepsilon}_x$) and Table 1 (or Table S1) can be used to determine the boundaries for ε_x (and x), given a certain level of confidence. It is very interesting to analyze two specific trends in Table 2. The first one is the relationship between $\bar{\varepsilon}_x$ and σ_a . There is a shift from negative bias values to positive ones with increasing values of σ_a . This might be explained by the behavior described in Fig. 3, as the curves starts slightly shifted towards negative values of ε_x and gradually move towards positive ones. The second aspect is the expected correlation between σ_a and S_x . From the original hypothesis in Eq. (17), it is supposed that the experimental errors are mostly due to catalyst fluctuations, so it is naturally expected that larger levels of σ_a generate larger S_x .

As an example on how to use the procedure, suppose one wants to find the confidence interval for \bar{x} . Assuming that the average conversion is equal to 0.80 and that the sample standard deviation is equal to 0.23 (assumed to be high on purpose, in order to magnify the size of the confidence region), the Gaussian distribution provides the 95% confidence interval of $0.34 < x < 1.26$, which is physically meaningless ($x > 1$). Using the proposed approach, instead, the first step is to search Table S2 for the row with $x = 0.80$ and, within it, for the sample standard deviation equal to 0.23. This corresponds to the row with $\sigma_a = 0.80$. With this value at hand, the next step is to search Table S1 for its portion corresponding to $x = 0.80$ and $\sigma_a = 0.80$ and finally find the values for the lower and upper boundaries of ε_x . In this particular case, they correspond to -0.65 and 0.18 (round values), respectively. Finally, these values are summed with $x = 0.80$ so that the 95% confidence interval becomes $0.15 < x < 0.98$, which is within the feasible experimental region, as one might expect from a sound statistical analysis of the available data. This sequence of steps is summarized in Fig. 4. This procedure could be replicated for S_x and for x in place of ε_x .

However, it is important to acknowledge that sample means and sample variances are also uncertain to a certain extent. In this case, distributions for means, standard deviations and single samples must be devised, similarly to the standard t-Student, F-Fisher and chi-square distributions for Gaussian distributions. In other words, it is necessary to derive the analogous kinetic t-Student, F-Fisher and chi-square distributions. It can be very laborious to compute these kinetic distributions analytically, as variables are correlated nonlinearly and because these distributions depend at least on three parameters: nominal conversion, catalyst activity variability and degrees of freedom. For this

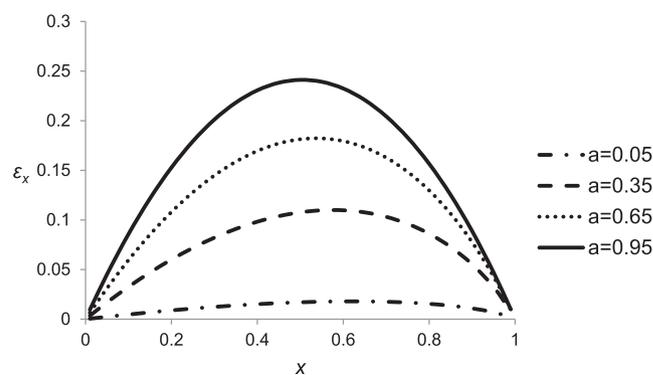


Fig. 2. Effect of catalyst activity and conversion on conversion fluctuations, in accordance with the proposed kinetic distribution.

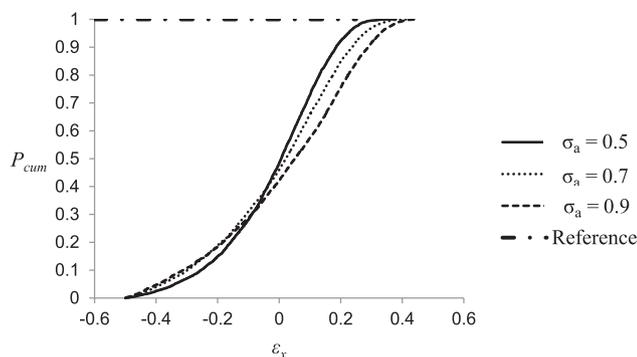


Fig. 3. Cumulative kinetic probability distribution of ε_x for different catalyst activity variabilities, for a reference conversion value of 0.5.

Table 1

Boundary values of ε_x as a function of catalytic activity variability (σ_a), as computed with the kinetic and Gaussian distributions, for several nominal conversions and confidence level of 95%.

| | $\sigma_a = 0.5$ | | $\sigma_a = 0.7$ | | $\sigma_a = 0.9$ | |
|-----------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| $x = 0.1$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ |
| | -0.0817 | 0.0876 | -0.0899 | 0.1238 | -0.0902 | 0.1528 |
| | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ |
| $x = 0.5$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ |
| | -0.3744 | 0.2496 | -0.4249 | 0.3191 | -0.4476 | 0.3602 |
| | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ |
| $x = 0.9$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ | $\varepsilon_{x\min}$ | $\varepsilon_{x\max}$ |
| | -0.542 | 0.0891 | -0.6771 | 0.0961 | -0.7208 | 0.0986 |
| | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ | $\varepsilon_{\text{normal}\min}$ | $\varepsilon_{\text{normal}\max}$ |
| | -0.337 | 0.337 | -0.4037 | 0.4037 | -0.4194 | 0.4194 |

Table 2

Relationship between catalytic activity deviation (σ_a) and first kinetic moments for conversion ($\bar{\varepsilon}_x$ and S_x) for a conversion of 50%.

| σ_a | $\bar{\varepsilon}_x$ | S_x |
|------------|-----------------------|-------|
| 0.1 | -0.00004 | 0.035 |
| 0.2 | -0.0044 | 0.071 |
| 0.3 | -0.0096 | 0.106 |
| 0.4 | -0.0094 | 0.138 |
| 0.5 | -0.010 | 0.167 |
| 0.6 | -0.013 | 0.183 |
| 0.7 | -0.0026 | 0.2 |
| 0.8 | 0.0093 | 0.213 |
| 0.9 | 0.020 | 0.222 |

reason, these distributions are computed here with help of numerical procedures, as shown in Fig. 5.

Fig. 5 illustrates the procedure that can be used to compute the sample kinetic distributions, which is analogous to the procedure shown in Fig. 1. Using the sample standard deviation as reference (as calculated with Eq. (25)), the reference variability of the catalyst activity (σ_a) can be defined with help of Table 2 (or Table S2). Then, NE sets of N random numbers (5000 random numbers were used here) are generated. For each of these sets, the values of a are calculated, using a reference distribution with zero mean (respecting the lower boundary of $a > -1$, as discussed in Section 3 and in Appendix A). One more time, a normal distribution was used here as reference without loss of generality. With the computed values of a , ε_x can be calculated with help of Eqs. (17) or (18). Finally, the first moments of the kinetic distribution are computed with the NE sampled values and the boundaries for \bar{x} and S_x can be defined with the desired confidence level. Although the procedure seems laborious, it can be easily performed in standard spreadsheets.

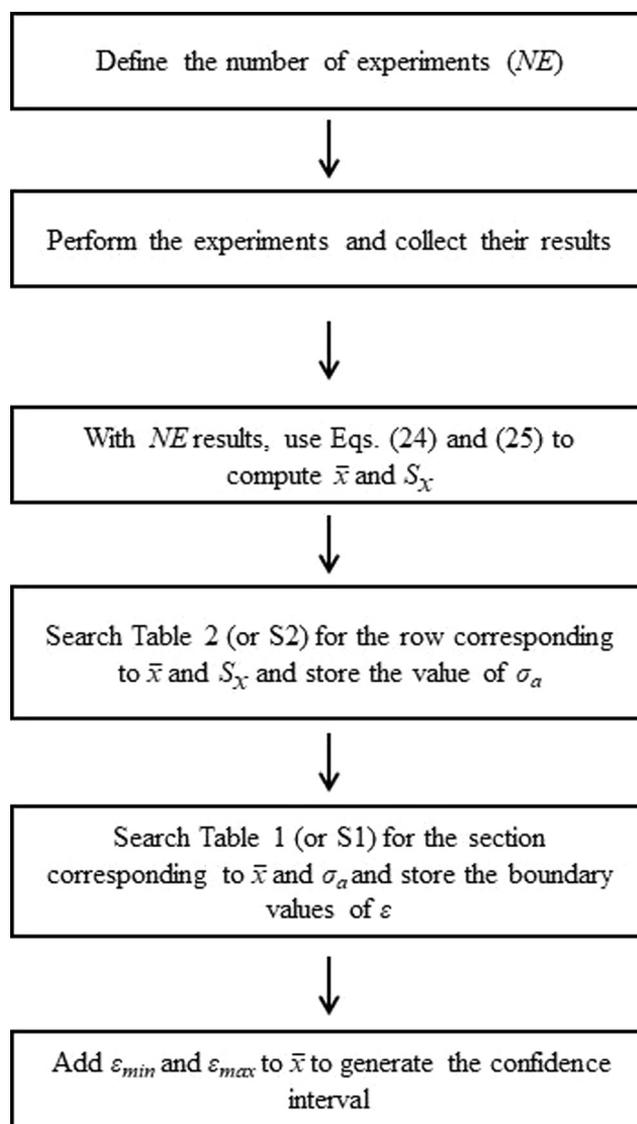


Fig. 4. Step-by-step instructions to apply the proposed methodology.

For illustrative purposes, Figs. 6 and 7 depict the behavior of the cumulative probability for the distributions of deviations of \bar{x} and S_x at three distinct levels of σ_a , for nominal conversion of 0.5 and $NE = 5$. For instance, for $\sigma_a = 0.5$ and confidence level of 95%, Figs. 6 and 7 may be used to obtain the upper and lower boundaries respectively of \bar{x} and S_x . The upper and lower boundary values for \bar{x} (this is done by selecting respectively $P_{ac} = 0.975$ and $P_{ac} = 0.025$ and then reading from Fig. 6 the value of ε_x) are 0.12 and -0.17 . With help of Table 2, the value of S_x corresponding to $x = 0.5$ and $\sigma_a = 0.5$ is 0.167 and the upper and lower boundary values, from Fig. 7, are respectively 0.27 and 0.055. Therefore, the confidence intervals become $0.33 < \mu_x < 0.62$ and $0.112 < \sigma_x < 0.437$. In order to use the proposed numerical approach, it is assumed that the conversion sample mean is equal to the real unknown conversion value, so that $x = \bar{x}$. Although this assumption can be regarded as a strong one, Table S2 shows that average conversion values are relatively insensitive to measurement fluctuations, so that this hypothesis is not thoroughly unreasonable.

5. Conclusion

The present manuscript presented a new family of non-normal error distributions to describe fluctuations of conversion values (and selectivities) in kinetic problems, named as the family of kinetic

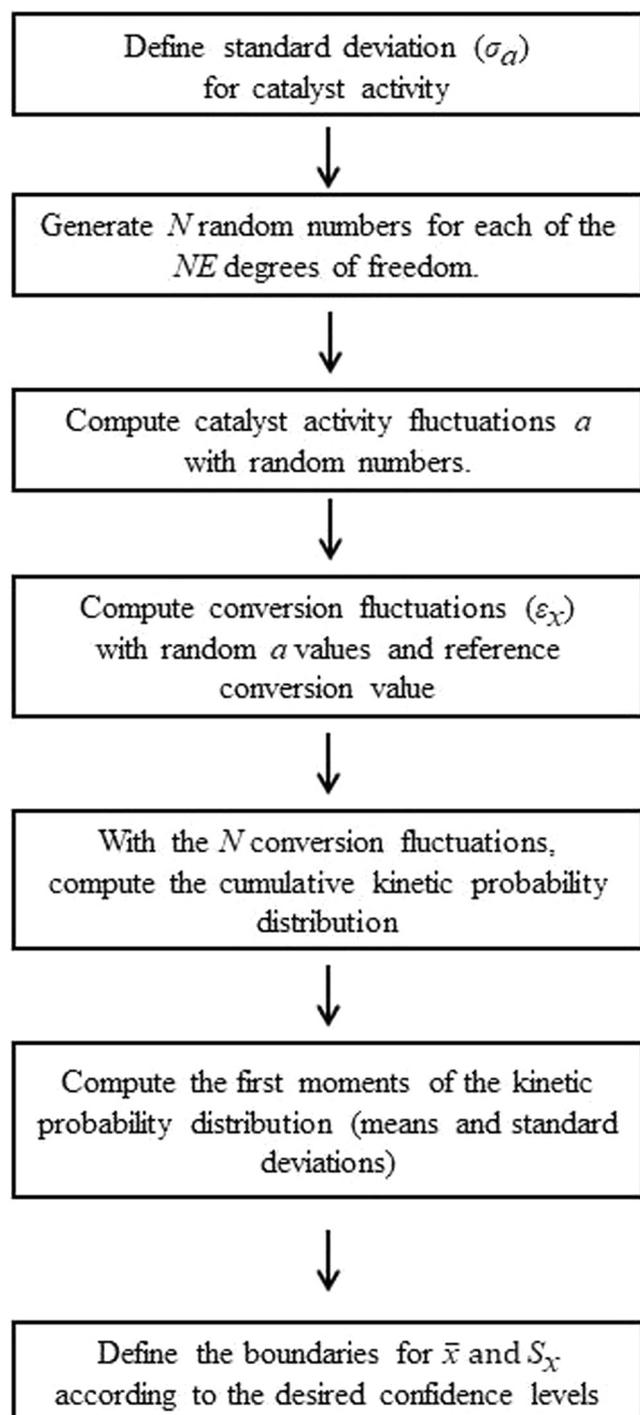


Fig. 5. Procedure for definition of the kinetic probability distribution for the mean and the standard deviation.

distributions. The derivation of the kinetic distributions considered a first-order reaction system, as analytical solutions can be easily devised in this case, but the proposed analyses can be extended to more involving kinetic systems in a straightforward manner with help of

Appendix A. Influence of the reference distribution selected to describe catalyst activity fluctuations

Eq. (17) can lead to interesting behaviors when the range of possible inputs for parameter a is enlarged. Fig. 7 shows how the responses change when a becomes negative. The function loses its characteristic parabolic-like profile at the threshold $a = -1$ and diverges to increasingly negative ε_x values. Observing the definition of parameter a in Eq. (16) ($a = \varepsilon_k/k'$), it becomes clear that a must be constrained to the interval $]-1, \infty [$. If $a \leq -1$, it means that the deviation ε_k is equal to or larger (in absolute terms) than k' , which means that the sum $(k' + \varepsilon_k)$ is negative. This, in turn, implies that the kinetics of the problem is reversed in respect to Eq. (1). Therefore, this lower limit for a should be respected during the generation of

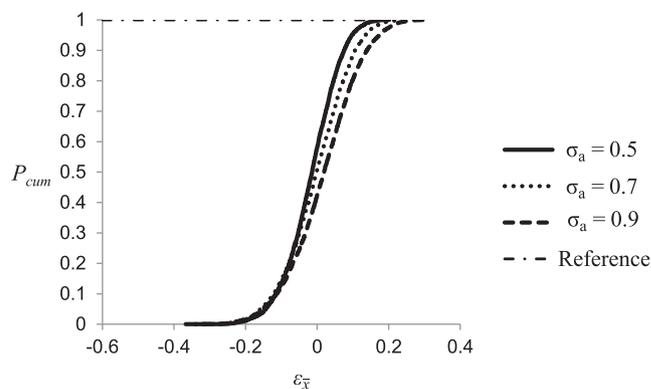


Fig. 6. Cumulative kinetic probability distribution for $\varepsilon_{\bar{x}}$ for nominal conversion of 0.50, $NE = 5$, and distinct catalyst activity variabilities.

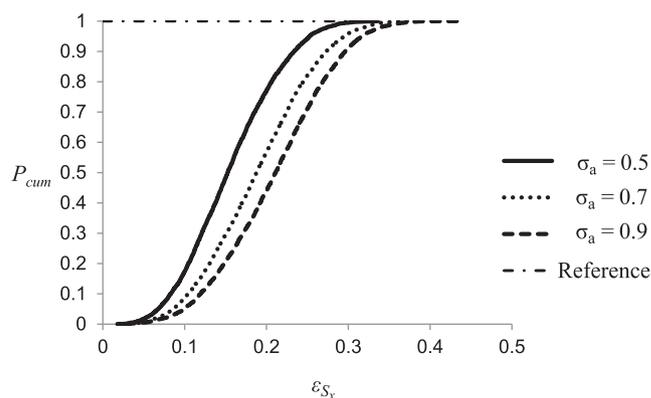


Fig. 7. Cumulative kinetic probability distribution for ε_{S_x} for nominal conversion of 0.50, $NE = 5$, and distinct catalyst activity variabilities.

numerical solutions provided by mass balance equations. Although the Gaussian distribution finds widespread use for a number of reasons, the fact is that confidence intervals computed with the Gaussian distribution can be physically incorrect in many kinetic problems, including conversions outside the $[0,1]$ range, for instance. As shown through a number of examples, the proposed kinetic distributions can always provide meaningful confidence intervals for conversions and selectivities. It is important to highlight that the kinetic distribution is intrinsically heteroscedastic and asymmetrical (differently from the Gaussian distribution, which is symmetrical and usually relies on constant variances for computations). Finally, a numerical procedure was proposed for computation of analogous t-Student, F-Fisher and chi-square distributions, using the kinetic distribution as reference, allowing for analyses of sample means and sample variances as obtained experimentally.

Acknowledgments

The authors thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for providing scholarships and for supporting this work.

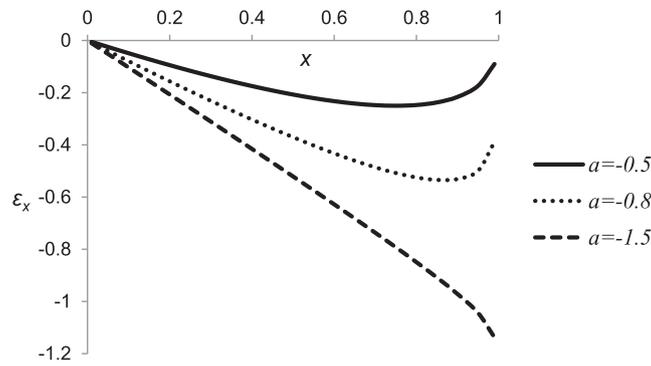


Fig. A1. Behavior of Eq. (17) for negative values of a .

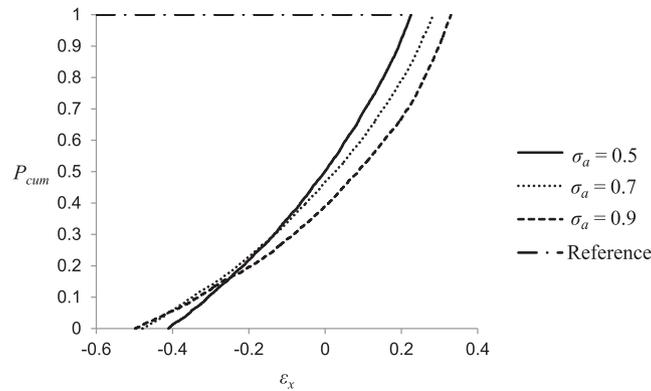


Fig. A2. Cumulative kinetic probability distribution of ϵ_x for different catalyst activity variabilities, for nominal conversion of 0.5, using the uniform distribution for computation of a .

random values in the numerical procedures of Fig. 1 and Fig. 4. This can be easily accomplished if the reference distribution used for generation of catalyst activities is defined in the feasible interval, such as the log-normal distribution and the uniform distribution.

Another feature that demands examination is the influence of the reference distribution selected to generate the catalyst activity values. While the normal distribution is very common and quite convenient to manipulate, it is certainly not the only option and should not be regarded as the most convenient reference *a priori*. Several well-known distributions could be used as references, such as the uniform, log-normal, exponential, binomial, among many others. In order to illustrate how the selection of a different reference distribution for catalyst activities can affect the outcome of the kinetic distribution, the uniform distribution is used for comparative purposes. In this case, the reference probability density function must be expressed as [11]:

$$f_z(z) = \begin{cases} \frac{1}{z_{max} - z_{min}}, & z_{min} < z < z_{max}, \\ 0, & otherwise \end{cases} \tag{A1}$$

whose mean and standard deviation are given respectively by:

$$\mu_z = \frac{1}{2}(z_{min} + z_{max}) \tag{A2}$$

$$\sigma_z = \frac{\sqrt{3}}{6}(z_{max} - z_{min}) \tag{A3}$$

Inserting Eq. (A1) into Eq. (19), and defining parameter a as the appropriate variable, it is possible to correlate a with ϵ_x in the form:

$$f_{\epsilon_x}(\epsilon_x) = \frac{1}{a_{max} - a_{min}} \frac{1}{[(1-x) - \epsilon_x] \ln(1-x)} \tag{A4}$$

Using the same procedure described in Fig. 1, the cumulative probability distributions for three levels of σ_a , at nominal conversion of 0.50, can be generated, as displayed in Fig. A1. Even though the curves of Figs. 3 and A2 are distinct, since they were generated with different reference distributions, it is interesting to note that the upper and lower boundaries predicted for ϵ_x , at 95% confidence level, are remarkably similar (as shown in Table A1). Given that the normal and uniform distributions are completely dissimilar, it might be conjectured that, to an acceptable degree of

Table A1

Boundary values of ϵ_x as a function of catalytic activity variability using a normal and a uniform distribution as references for computation of a , for nominal conversion of 0.5 and confidence level of 95%.

| $\sigma_a = 0.5$ | | $\sigma_a = 0.7$ | | $\sigma_a = 0.9$ | |
|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| $\epsilon_{x_{min}}$ | $\epsilon_{x_{max}}$ | $\epsilon_{x_{min}}$ | $\epsilon_{x_{max}}$ | $\epsilon_{x_{min}}$ | $\epsilon_{x_{max}}$ |
| -0.374 | 0.250 | -0.425 | 0.319 | -0.448 | 0.360 |
| ϵ_{unimin} | ϵ_{unimax} | ϵ_{unimin} | ϵ_{unimax} | ϵ_{unimin} | ϵ_{unimax} |
| -0.382 | 0.212 | -0.441 | 0.276 | -0.453 | 0.322 |

numerical rigor, the distribution selected to represent the fluctuations of catalyst activity exerts relatively low impact on the final calculations performed with the proposed kinetic distribution.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cej.2017.09.076>.

References

- [1] G.F. Froment, K.B. Bischoff, *Chemical Reactor Analysis and Design*, 1st ed., John Wiley & Sons, New York, 1979.
- [2] M. Schwaab, J.C. Pinto, *Análise de Dados Experimentais I: Fundamentos de Estatística e Estimação de Parâmetros*, 1st ed., E-Papers, Rio de Janeiro, 2007.
- [3] K. Raji, S. Thomas, C.B. Sobhan, A chemical kinetic model for chemical vapor deposition of carbon nanotubes, *Appl. Surf. Sci.* 257 (2011) 10562–10570.
- [4] O.A. Gutiérrez, M. Chavez, E. Lissi, A theoretical approach to some analytical properties of heterogeneous enzymatic assays, *Anal. Chem.* 76 (2004) 2664–2668.
- [5] F. Yin, H. Chen, G. Xu, G. Wang, Y. Xu, A detailed kinetic model for the hydrothermal decomposition process of sewage sludge, *Bioresour. Technol.* 198 (2015) 351–357.
- [6] M. Schwaab, E.C. Biscaia Jr., J.L. Monteiro, J.C. Pinto, Nonlinear parameter estimation through particle swarm optimization, *Chem. Eng. Sci.* 63 (2008) 1542–1552.
- [7] M. Schwaab, J.C. Pinto, Optimum reference temperature for reparameterization of the Arrhenius equation. Part I: Problems involving one kinetic constant, *Chem. Eng. Sci.* 62 (2007) 2750–2764.
- [8] M. Schwaab, L.P. Lemos, J.C. Pinto, Optimum reference temperature for reparameterization of the Arrhenius equation. Part 2: Problems involving multiple reparameterizations, *Chem. Eng. Sci.* 63 (2008) 2895–2906.
- [9] M. Schwaab, J.C. Pinto, Optimum reparameterization of power function models, *Chem. Eng. Sci.* 63 (2008) 4631–4635.
- [10] A.L. Alberton, M. Schwaab, M. Schmal, J.C. Pinto, Experimental errors in kinetic tests and its influence on the precision of estimated parameters. Part I - Analysis of first-order reactions, *Chem. Eng. J.* 155 (2009) 816–823.
- [11] D.C. Montgomery, G.C. Runger, *Applied Statistics and Probability for Engineers*, 3rd ed., John Wiley and Sons Inc, New York, 2002.
- [12] B.A. Ogunnaike, W.H. Ray, *Process Dynamics, Modeling, and Control*, 1st ed., Oxford University Press, New York, 1994.
- [13] T.C. Castonguay, F. Weng, Kinetic Monte Carlo modeling of chemical reactions coupled with heat transfer, *J. Chem. Phys.* 128 (2008) 1247061–1247067.
- [14] L.P. Oliveira, J.J. Verstraete, M. Kolb, A Monte Carlo modeling methodology for the simulation of hydrotreating processes, *Chem. Eng. J.* 207 (2012) 94–102.
- [15] X.Q. Zhang, R.A. van Santen, A.P.J. Jansen, Kinetic Monte Carlo modeling of silicate oligomerization and early gelation, *Phys. Chem. Chem. Phys.* 14 (2012) 11969–11973.