On Extreme Concentrations in Chemical Reaction Networks with Incomplete Measurements

Rafael Blanquero,[†] Emilio Carrizosa,[†] Oana Chis,[‡] Noemí Esteban,[§] Asunción Jiménez-Cordero,[*,†] José Francisco Rodríguez,[‖] and M. Remedios Sillero-Denamiel[⊥]

[†]Departamento de Estadística e Investigación Operativa, and [⊥]Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Universidad de Sevilla, 41012, Seville, Spain

[‡]Instituto Tecnológico de Matemática Industrial (ITMATI), 15782, Santiago de Compostela, Spain

[§]Departamento de Matemática Aplicada, Universidad de Santiago de Compostela, 15782, Santiago de Compostela, Spain

[‖]Repsol Technology Center, 28935, Móstoles, Madrid, Spain

ABSTRACT: A fundamental problem in the analysis of chemical reactions networks consists of identifying concentration values along time or in steady state which are coherent with the experimental concentration data available. When concentration measurements are incomplete, either because information is missing about the concentration of a species at a particular time instant, or even there is no information at all on the concentration of a species, then the problem becomes ill-defined, and then different concentration curves are compatible with existing data. In this paper we address the problem of finding the extreme (highest and lowest) concentrations under incomplete data measurements; as a byproduct of our approach, the model parameters associated with such extreme concentrations are obtained. These extreme concentrations provide valuable information on the impact that incomplete measurements have on the theoretical reconstruction of concentrations from experimental data. To obtain such concentrations range, mathematical optimization problems are formulated, solvable by a variety of global optimization approaches, such as, for example, the stochastic global optimization method suggested.
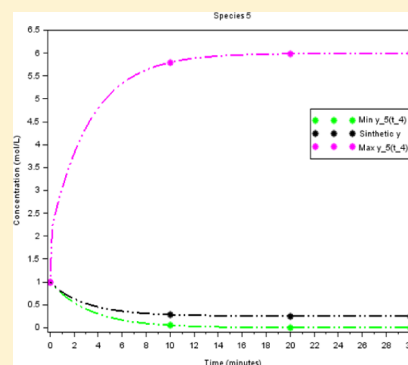
1. INTRODUCTION

In a chemical reaction network, a set of N species, $E_1, ..., E_N$, react in L chemical reactions. The vector $\mathbf{y}$ of concentrations of the different species $E_1, ..., E_N$ evolves along time according to some kinetic law,

$$\begin{cases} \dot{\mathbf{y}}(t) = \mathbf{F}(t, \mathbf{y}(t), \mathbf{u}(t), \boldsymbol{\theta}), \ t \in [0, T] \\ \mathbf{y}(0) = \mathbf{y}_0 \end{cases} \tag{1}$$

where $\mathbf{y}_0 \in \mathbb{R}^N$ is the vector of initial concentrations, t is the independent variable that denotes the time, $\mathbf{u}$ is the vector of inputs, and $\mathbf{F}$ gives the expression of the right-hand side of the ordinary differential equation representing the kinetic law. Finally, $\boldsymbol{\theta}$ is the vector of unknown parameters which takes values in some set Θ. In what follows, the dot over the function, as in $\dot{\mathbf{y}}(t)$, denotes the derivative of the function with respect to time, and the bold letters, as $\boldsymbol{\theta}$ or $\mathbf{y}$, are used to represent vectors. While eq 1 is used to study the transient state of the reactions network, the concentrations at the steady state are obtained by solving the system of nonlinear equations

$$\mathbf{F}(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) = 0 \tag{2}$$

Inferring the value of $\boldsymbol{\theta}$ from experimental data is a challenge,[3,6,7,22] since a difficult mathematical optimization problem is to be solved to find the value of $\boldsymbol{\theta}$ yielding the best fit,.[11,13,14,20,23,24] Such a problem may be ill-posed,[14] since there

may exist different vectors $\boldsymbol{\theta}$ which fit well the experimental data and at the same time they satisfy the constraints given by eq 1 in the transient state or by eq 2 in the steady state. Consequently, some methods, such as ensemble modeling,[26] find the distribution of parameters rather than particular values, and therefore the behavior of a population of parameters which is consistent with the information on the system is described. Regularization,[13,14] is a popular approach to infer one possible solution $\boldsymbol{\theta}$. However, regularization is not fully satisfactory; indeed, penalty coefficients are associated with the different parameters to be inferred, which may be in different scales (e.g., very different orders of magnitude of reaction rates). This makes it hard to properly tune the regularization coefficients, and thus, the solution $\boldsymbol{\theta}$ obtained by solving the regularized problem may be far from the actual one.

Incremental methods, such as the one in ref 21 are also used to infer a solution $\boldsymbol{\theta}$. The methodology in ref 21 reduces the parameter space, making then lower the computational cost associated with the parameter inference problem. However, this approach can only be applied for reaction networks satisfying certain assumptions; for example, the number of reactions is bigger than the number of species and particular types of kinetic

law where $\mathbf{F}(t, \mathbf{y}(t), \mathbf{u}(t), \boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})\delta(t, \mathbf{y}(t), \mathbf{u}(t), \boldsymbol{\theta})$, with $\mathbf{A}(\boldsymbol{\theta})$ being the stoichiometric matrix, and δ the reaction rate flux. For this reason, it would be very helpful to explore whether the system is structurally identifiable, that is, if we could know a priori if the parameters would be univocally identified. The drawback of the identifiability method is the complexity of the model (number of parameters, high degree of nonlinearity). See ref 10 for a critical comparison of existing techniques and ref 9 for a software tool. Observe that when studying structural identifiability, we consider only the system dynamics, the observations, and the stimuli of the model. Structural identifiability regards the possibility of giving unique values to unknown model parameters from the available measurements, assuming perfect experimental data (i.e., noise-free and continuous in time).[27] Nevertheless, according to ref 25, in order to determine r parameters of an identifiable system, one requires $2r + 1$ properly chosen measurements. Such number of measurements may be high and impossible to have in some cases. See ref 10 for a more detailed structural identifiability definition and ref 17 for methods to test structural identifiability. In ref 27 the authors have done a comparison of the existing methods and also have presented a new method based on the generating series approach. On the other hand, practical identifiability is related with the computation of numerical parameter estimates and is based on experimental data and noise. See ref 10 and its references for a deeper analysis.

Hence, the incomplete measurements of the experimental data may make meaningless the search of $\boldsymbol{\theta}$, since there may exist multiple values of $\boldsymbol{\theta}$ that explain the experimental data, and therefore there are different solutions $\mathbf{y}(t)$ which are coherent with the experimental values. In what follows, we mean by incomplete measurement two different things: on the one hand, an incomplete measurement implies the lack of information about the concentration of a species in some time instant. On the other hand, it indicates that the concentration of a species is unknown for all the time instants (except the initial one). Two dramatic consequences should be highlighted when incomplete measurements appear. First, some approaches such as the extent-based method,[1,2,15] are not directly applicable owing to the lack of experimental values for all the species. Second, even if the approach can handle incomplete measurements, the so-obtained output may be inaccurate.[2]

Several approaches have been proposed in the literature to deal with incomplete measurements. One of the most popular techniques is inputation,[12] which may yield wrong results if the inputted values are not close enough to the actual (incomplete) measurements. Alternatively, the approach proposed in refs 4 and 5 reconstructs sometimes the missing concentrations values from the stoichiometry information, but it is not always able to perform such reconstruction. As an illustration, let us focus on the steady state of the reactions network in eq 3 where six species, E_1, ..., E_6 and a catalyst, Cat, interact in a continuous stirred tank reactor.[2]

$$
\begin{aligned}
E_1 + E_2 &\xrightarrow{\text{Cat}} E_3 \\
E_2 + E_2 &\xrightarrow{\text{Cat}} E_4 \\
E_2 &\longrightarrow E_5 \\
E_3 + E_2 &\xrightarrow{\text{Cat}} E_6
\end{aligned}
\tag{3}
$$

The six-dimensional vector of concentrations, $\mathbf{y}$, is the solution of the following system of equations:

$$
\begin{aligned}
F_1(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) &= -\theta_1 y_1 y_2 y_{\text{Cat}} - u_{\text{out}} y_1 = 0 \\
F_2(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) &= -\theta_1 y_1 y_2 y_{\text{Cat}} - 2\theta_2 y_2^2 y_{\text{Cat}} - \theta_3 y_2 - \theta_4 y_3 y_2 y_{\text{Cat}} \\
&\quad + W_2 u_{\text{in}} - u_{\text{out}} y_2 = 0 \\
F_3(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) &= \theta_1 y_1 y_2 y_{\text{Cat}} - \theta_4 y_3 y_2 y_{\text{Cat}} - u_{\text{out}} y_3 = 0 \\
F_4(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) &= \theta_2 y_2^2 y_{\text{Cat}} - u_{\text{out}} y_4 = 0 \\
F_5(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) &= \theta_3 y_2 - u_{\text{out}} y_5 = 0 \\
F_6(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) &= \theta_4 y_3 y_2 y_{\text{Cat}} - u_{\text{out}} y_6 = 0
\end{aligned}
\tag{4}
$$

In eq 4, the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$ plays the role of the vector of rate constants. The magnitudes of θ_1, θ_2, and θ_4 are $\text{L}^2\,\text{mol}^{-2}\,\text{min}^{-1}$, and the magnitude of θ_3 is min^{-1}; y_i, for $i = 1, ..., 6$ are the components of the vector $\mathbf{y}$ and designate the steady states of the concentrations measured in mol/L of the species E_i, while y_{Cat} denotes the concentration of the catalyst Cat and has the value $y_{\text{Cat}} = 0.5$ mol/L. Furthermore, $u_{\text{in}} = 0.3$ L/min and $u_{\text{out}} = 0.3$ L/min are the inlet and outlet mass flow rates, respectively. Finally, W_2 is the second component of the inlet composition vector $\mathbf{W} = (0, 6, 0, 0, 0, 0)$ mol/L.

Let us assume we are given the experimental data in Table 1, in which only experimental measurements for species E_2, E_4, and E_6

Table 1. Experimental Concentrations in the Steady State of the Example with Model (eq 4)

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn		2.6815		1.5341		$<10^{-8}$

are provided, while data on E_1, E_3, and E_5 are missing. The concentration of E_6 is expressed as $<10^{-8}$ since the actual value is very close to zero, and therefore the particular value of the concentration will be strongly dependent on the accuracy of the measuring machine. This acceptance does not imply that the species with low concentrations are treated as missing, but that it is impossible to know a priori the exact value, since it depends on the machine used.

If the approach of refs 4 and 5 is used, it is seen that none of the incomplete measurements of the species can be reconstructed. In fact, both parameter values $\boldsymbol{\theta}$ in eq 5 and eq 6 give concentrations values coherent with the observed ones, as given in Table 2.

$$
\boldsymbol{\theta} = (0.053, 0.1280, 0.0280, 0.0001)
\tag{5}
$$

$$
\boldsymbol{\theta} = (0.4182, 0.1280, 0.0280, 1.5428)
\tag{6}
$$

Table 2. Theoretical Concentrations of the Example with Model eq 4 and with $\boldsymbol{\theta}$ in eq 5 or eq 6

	E_1	E_2	E_3	E_4	E_5	E_6
y for $\boldsymbol{\theta}$ in eq 5 or eq 6	$<10^{-8}$	2.6815	$<10^{-8}$	1.5341	0.2502	$<10^{-8}$

In this paper, instead of inferring one possible value for $\boldsymbol{\theta}$, which may be right or wrong, we seek the extreme concentrations (highest and lowest values) of the incomplete measurements which are in agreement with the experimental data. As a byproduct, the parameters associated $\boldsymbol{\theta}$ are also obtained. To do this, global optimization problems are to be solved. Two types of methods are available to cope with such problems, namely, deterministic and stochastic global optimization techniques. Deterministic methods, as interval analysis based branch and

bound,[23,24] do find the global optimum in finite time, at the expense of high computational burden and strong implementation effort. Therefore, their usefulness is mainly limited to offline applications and small reactions networks. In contrast, stochastic methods, such as variable neighborhood search (VNS),[19] are much easier to implement and faster to execute.

The main contribution of this paper is the analysis of incomplete measurements by means of the search of the extreme concentrations. For simplicity, throughout the paper we have chosen to solve the optimization problems found with a stochastic optimization tool, namely VNS.

The remainder of this paper is structured as follows. In section 2 we present our approach in the simplest case in which the problem is studied in the steady state. Then, a nontrivial extension, including the analysis for problems in transient state, is described in section 3. Some final conclusions and possible research lines are outlined in section 4.

2. PROBLEM STATEMENT. THE STEADY STATE

In this section the steady state of a chemical reaction network is considered, and the extreme concentrations (i.e., those yielding highest or lowest concentration values) are sought when measurements are incomplete. When multiple steady states exist, the goal is the same, since we will find the maximum and minimum values of the unknown concentrations for all the possible steady states, and their associated parameters.

2.1. Mathematical Model. Let us split the index set $\{1, ..., N\}$ into O and M which correspond respectively to the set of species E_i for which the steady-state concentrations are observed and are missing. Accordingly, the N-dimensional vector $\hat{\mathbf{y}}$ of steady-state concentrations is split into two blocks,

$$\hat{\mathbf{y}} = (\widehat{\mathbf{y}^{obs}}|\widehat{\mathbf{y}^{mis}})$$

where $\widehat{\mathbf{y}^{obs}}$ represents the experimental values of the observed species, E_i, $i \in O$, while $\widehat{\mathbf{y}^{mis}}$ denotes the values of the missing species, E_i, $i \in M$.

For each species E_i, $i \in M$, the concentration value y_i is uncertain. Let us bound the degree of uncertainty by imposing such unknown y_i to belong to a certain set $\mathcal{Y}_i \subset [0, \infty)$ If no information is provided, one can set $\mathcal{Y}_i = [0, \infty)$

With this notation at hand, we are in position to formally state our goal: we seek, for each species E_i, $i \in M$ the lowest and highest values $\underline{y}_i$ and $\overline{y}_i$ which are compatible with the steady-state eqs 2; such extreme concentrations will be associated with some $\theta \in \Theta$, denoted respectively as $\underline{\theta}_i$ and $\overline{\theta}_i$ referred as extreme parameters since they yield the extreme concentrations.

The extreme concentrations $\underline{y}_i$, $\overline{y}_i$ and extreme parameters $\underline{\theta}_i$ and $\overline{\theta}_i$ are obtained respectively as optimal value and optimal solution of the following optimization problems, for all $i \in M$:

$$\begin{cases} \underline{\theta}_i = \arg\min & y_i \\ \text{s.t.} \ \mathbf{F}(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) = \mathbf{0} \\ \quad |y_j - \widehat{y_j^{obs}}| \le \varepsilon, \quad j \in O \\ \quad y_l \in \mathcal{Y}_l, \quad l \in M \\ \quad \boldsymbol{\theta} \in \Theta \end{cases} \quad (7)$$

$$\begin{cases} \overline{\theta}_i = \arg\max & y_i \\ \text{s.t.} \ \mathbf{F}(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) = \mathbf{0} \\ \quad |y_j - \widehat{y_j^{obs}}| \le \varepsilon, \quad j \in O \\ \quad y_l \in \mathcal{Y}_l, \quad l \in M \\ \quad \boldsymbol{\theta} \in \Theta \end{cases} \quad (8)$$

Observe that we are allowing the model to yield, for $j \in O$ values y_j differing from the observed $\widehat{y_j^{obs}}$ in at most some $\varepsilon \ge 0$, allowing one to accommodate some uncertainty in the measurement of such observed species E_j.

The optimization problems eq 7 and eq 8 have a simple objective function, but the constraints are highly nonlinear, yielding a difficult nonconvex optimization problem. Hence, standard local-search optimization techniques may fail to properly identify the optimal solutions. We discuss in section 2.2 how the global optima can be obtained by using the well-known metaheuristics, namely variable neighborhood search, in short VNS.[8,18,19]

2.2. Inferring Extreme Concentrations by Stochastic Global Optimization. Since the approach is the same for Problems 7 and 8, we focus in what follows on Problem 7 for the sake of simplicity. The idea of VNS is simple: the algorithm starts by seeking a local optima from an initial point; once a local optimum has been obtained, the solution is (slightly) perturbed and then a new local search is performed from the perturbed solution, with the hope of attaining a better local optimum. If the so-obtained solution does not improve the incumbent, then a new search is performed, starting now from a point obtained by introducing a stronger perturbation in the best solution obtained so far. Two elements are critical in the process. First, how the process is initialized, that is, how the first local-search is started. Second, how solutions are perturbed. To construct a starting solution, the extent-based method is suggested.[1] Perturbations of a feasible solution $\boldsymbol{\theta}$ are obtained by randomly generating a point in a neighborhood

$$\mathcal{N}_R(\boldsymbol{\theta}) = \prod_i [\theta_i - R\gamma, \theta_i + R\gamma]$$

for some integer R, where γ is a scaling parameter.

Afterward, in a general iteration of the VNS process, we randomly select a point $\boldsymbol{\theta}^0$ from the neighborhood $\mathcal{N}_R(\boldsymbol{\theta}^{opt})$, $\boldsymbol{\theta}^{opt}$ being the best solution so far obtained. Next, Problem 7 is solved by means of a local search method, using $\boldsymbol{\theta}^0$ as initial solution, yielding the vector $\boldsymbol{\theta}^*$. If $\boldsymbol{\theta}^*$ yields an objective value better than $\boldsymbol{\theta}^{opt}$, then $\boldsymbol{\theta}^{opt}$ is updated. Otherwise, the neighborhood around $\boldsymbol{\theta}^{opt}$ used to perturb solutions is enlarged, that is, we will perturb $\boldsymbol{\theta}^{opt}$ by generating some random $\boldsymbol{\theta} \in \mathcal{N}_{R+1}(\boldsymbol{\theta}^{opt})$. This process is repeated until a stopping criteria is fulfilled (time limit or iterations limit). A pseudocode of the optimization process is shown in Algorithm 1.

Algorithm 1: VNS pseudocode

Initialization
- Define R_{max}.
- Solve Problem (7) with the Extent-based Method, obtaining the solution $\boldsymbol{\theta}^{opt}$.

repeat
  $R := 1$
  while $R \le R_{max}$ **do**
    • Select a random initial solution $\boldsymbol{\theta}^0$ from $\mathcal{N}_R(\boldsymbol{\theta}^{opt})$.
    • Solve Problem (7) with a local search technique, starting from $\boldsymbol{\theta}^0$. The solution so obtained is denoted by $\boldsymbol{\theta}^*$.
    if $\boldsymbol{\theta}^*$ *is better than* $\boldsymbol{\theta}^{opt}$ **then**
      $\boldsymbol{\theta}^{opt} := \boldsymbol{\theta}^*$
      $R := 1$
    else
      $R := R + 1$
    end
  end
until *stopping condition is reached;*

2.3. Examples. In this section, four examples are provided to illustrate our approach. The examples of sections 2.3.1, 2.3.2, and 2.3.3 correspond to the data from ref 2, while in section 2.3.4 a realistic and industrial example coming from the petrochemicals company REPSOL is studied. In all the examples we have analyzed the results of our methodology by using, first, concentrations without noise and, second, concentrations corrupted with additive zero-mean Gaussian noise with a standard deviation of 1%. This model of noise has been taken from ref 2. Our VNS algorithm has been coded in Fortran and compiled using Intel©Fortran Compiler XE 12.0. Executions were carried out on an Intel Core i7 computer with 16.00 Gb of RAM memory at 2.6 GHz, running Windows 8. In the implementation of the VNS, the number R_{max} of different radii was set equal to 10 and the value of γ is 5 for sections 2.3.1, 2.3.2, and 2.3.3, while $R_{max} = 20$ and $\gamma = 500$ for section 2.3.4. A sequential equality constrained quadratic programming method is used as a local-search routine by means of the functions NNLPF and the nonlinear systems of equations are solved using a modified Powell hybrid algorithm through the function NEQNF. Both functions are available at the IMSL Fortran Numerical Library.

Furthermore, the tolerance ε in Problems 7 and 8 is set to 10^{-7} in sections 2.3.1, 2.3.2, and 2.3.3, while it is $\varepsilon = 10^{-2}$ in section 2.3.4 because the problems turned out to be much harder, as a solution with tolerance 10^{-7} was not found. In sections 2.3.1, 2.3.2, and 2.3.3, the set Θ is the cartesian product of intervals of the form $[10^{-10}, +\infty)$. Nevertheless, in section 2.3.4, the decision variables $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{Ea})$, as defined in eq 11, is allowed to take values, $\mathbf{a} \in [10^{-10}, 10^{15}]$ and $\mathbf{Ea} \in [10^{-10}, 10^{10}]$. Note that the admissible values of the parameters $\mathbf{a}$ and $\mathbf{Ea}$ should be understood componentwise. Finally, the sets where the theoretical concentrations of the incomplete species take values, $\mathcal{Y}_l$, $l \in \mathcal{M}$, are chosen to be $[10^{-7}, +\infty)$ in all the examples. As a preprocessing step, we have determined which are the species that can be identified by using the methodology in refs 4 and 5 We recall that in the steady state, this method is able to identify which species can be reconstructed, but it cannot reconstruct them, since at least one value per species is needed, and this information is not available in the steady state. We remark that we are not comparing our approach with the one in refs 4 and 5. Indeed, the method in refs 4 and 5 provides a powerful tool to reconstruct some incomplete measurements, but it may fail to reconstruct all of them, as mentioned above. In such cases, our method is able to find a solution coherent with the available information. Hence, we are not competing but complementing refs 4 and 5.
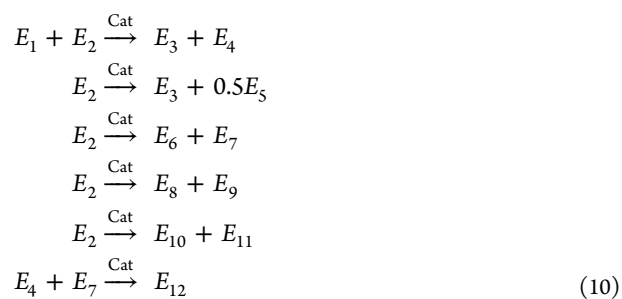
The experimental concentrations of all the species of sections 2.3.1, 2.3.2, and 2.3.3 have been generated using the parameters stated in ref 2 and given in eq 5, that is, $\boldsymbol{\theta}$ in eq 9, named the true $\boldsymbol{\theta}$ in what follows. These concentrations without noise are shown in Table 2.

$$\boldsymbol{\theta} = (0.0530, 0.1280, 0.0280, 0.0001) \tag{9}$$

On the other hand, the concentrations with noise are shown in Table 3.

The example of section 2.3.4 comes from the petrochemicals company REPSOL. It consists of the steady state of the reaction network in eq 10 where 12 species, $E_1, ..., E_{12}$ and a catalyst, Cat, react in a continuous stirred tank reactor:

$$
\begin{aligned}
E_1 + E_2 &\xrightarrow{\text{Cat}} E_3 + E_4 \\
E_2 &\xrightarrow{\text{Cat}} E_3 + 0.5E_5 \\
E_2 &\xrightarrow{\text{Cat}} E_6 + E_7 \\
E_2 &\xrightarrow{\text{Cat}} E_8 + E_9 \\
E_2 &\xrightarrow{\text{Cat}} E_{10} + E_{11} \\
E_4 + E_7 &\xrightarrow{\text{Cat}} E_{12}
\end{aligned}
\tag{10}
$$

The 12-dimensional vector of concentrations of the steady state, $\mathbf{y}$, obeys the following system of nonlinear equations:

$$F_1(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = -a_1 \exp\left(\frac{-Ea_1}{R \times \text{Temp}}\right) y_1 y_2 y_{\text{Cat}} + \frac{1}{V}(W_1 u_{\text{in}} - u_{\text{out}} y_1) = 0$$

$$F_2(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = -a_1 \exp\left(\frac{-Ea_1}{R \times \text{Temp}}\right) y_1 y_2 y_{\text{Cat}}$$
$$- a_2 \exp\left(\frac{-Ea_2}{R \times \text{Temp}}\right) y_2^2 y_{\text{Cat}}^{0.5} - a_3 \exp\left(\frac{-Ea_3}{R \times \text{Temp}}\right) y_2^2 y_{\text{Cat}}$$
$$- a_4 \exp\left(\frac{-Ea_4}{R \times \text{Temp}}\right) y_2 y_{\text{Cat}}^{1.1} - a_5 \exp\left(\frac{-Ea_5}{R \times \text{Temp}}\right) y_2 y_{\text{Cat}}^{0.5}$$
$$+ \frac{1}{V}(W_2 u_{\text{in}} - u_{\text{out}} y_2) = 0$$

$$F_3(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_1 \exp\left(\frac{-Ea_1}{R \times \text{Temp}}\right) y_1 y_2 y_{\text{Cat}}$$
$$+ a_2 \exp\left(\frac{-Ea_2}{R \times \text{Temp}}\right) y_2^2 y_{\text{Cat}}^{0.5} + \frac{1}{V}(W_3 u_{\text{in}} - u_{\text{out}} y_3) = 0$$

$$F_4(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_1 \exp\left(\frac{-Ea_1}{R \times \text{Temp}}\right) y_1 y_2 y_{\text{Cat}}$$
$$- a_6 \exp\left(\frac{-Ea_6}{R \times \text{Temp}}\right) y_4 y_7 y_{\text{Cat}}^{1.2} + \frac{1}{V}(W_4 u_{\text{in}} - u_{\text{out}} y_4) = 0$$

$$F_5(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = 0.5 a_2 \exp\left(\frac{-Ea_2}{R \times \text{Temp}}\right) y_2^2 y_{\text{Cat}}^{0.5} + \frac{1}{V}(W_5 u_{\text{in}} - u_{\text{out}} y_5) = 0$$

$$F_6(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_3 \exp\left(\frac{-Ea_3}{R \times \text{Temp}}\right) y_2^2 y_{\text{Cat}} + \frac{1}{V}(W_6 u_{\text{in}} - u_{\text{out}} y_6) = 0$$

$$F_7(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_3 \exp\left(\frac{-Ea_3}{R \times \text{Temp}}\right) y_2^2 y_{\text{Cat}}$$
$$- a_6 \exp\left(\frac{-Ea_6}{R \times \text{Temp}}\right) y_4 y_7 y_{\text{Cat}}^{1.2} + \frac{1}{V}(W_7 u_{\text{in}} - u_{\text{out}} y_7) = 0$$

$$F_8(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_4 \exp\left(\frac{-Ea_4}{R \times \text{Temp}}\right) y_2 y_{\text{Cat}}^{1.1} + \frac{1}{V}(W_8 u_{\text{in}} - u_{\text{out}} y_8) = 0$$

$$F_9(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_4 \exp\left(\frac{-Ea_4}{R \times \text{Temp}}\right) y_2 y_{\text{Cat}}^{1.1} + \frac{1}{V}(W_9 u_{\text{in}} - u_{\text{out}} y_9) = 0$$

$$F_{10}(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_5 \exp\left(\frac{-Ea_5}{R \times \text{Temp}}\right) y_2 y_{\text{Cat}}^{0.5} + \frac{1}{V}(W_{10} u_{\text{in}} - u_{\text{out}} y_{10}) = 0$$

$$F_{11}(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_5 \exp\left(\frac{-Ea_5}{R \times \text{Temp}}\right) y_2 y_{\text{Cat}}^{0.5} + \frac{1}{V}(W_{11} u_{\text{in}} - u_{\text{out}} y_{11}) = 0$$

$$F_{12}(\mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{Ea}) = a_6 \exp\left(\frac{-Ea_6}{R \times \text{Temp}}\right) y_4 y_7 y_{\text{Cat}}^{1.2}$$
$$+ \frac{1}{V}(W_{12} u_{\text{in}} - u_{\text{out}} y_{12}) = 0$$

$$\tag{11}$$

where the parameter vector has two different subvectors, $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{Ea})$, $\mathbf{a} = (a_1, ..., a_6)$ and $\mathbf{Ea} = (Ea_1, ..., Ea_6)$ denote, respectively, the prefactor and the activation energy of the

Table 3. Experimental Concentrations with Noise of Sections 2.3.1, 2.3.2, and 2.3.3

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn	$<10^{-8}$	2.6974	$<10^{-8}$	1.5431	0.2460	$<10^{-8}$

Table 4. Experimental Concentrations of Example in section 2.3.4

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn	0.1791	0.0329	0.9587	0.8172	0.0689	0.0081
	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
expt concn	0.0045	0.0001	0.0001	0.2333×10^{-8}	0.2333×10^{-8}	0.0035

Table 5. Experimental Concentrations with Noise of Example in section 2.3.4

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn	0.1755	0.0331	0.9495	0.8221	0.0678	0.0080
	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
expt concn	0.0044	0.0001	0.0001	0.2350×10^{-8}	0.2325×10^{-8}	0.0035

Arrhenius law.[16] The magnitude of $\mathbf{a}$ is L/mol·s and the magnitude of $\mathbf{Ea}$ is J/mol. The steady states of the concentrations y_i, $i = 1, ..., 12$ are measured in mol/L, while $y_{Cat} = 0.001$ mol/L is the concentration of the catalyst Cat. The temperature Temp = 383 K, and $R = 8.314$ J/mol·K denotes the universal gas constant. Moreover, the volume of the tank vessel is $V = 10$ L and $u_{in} = 0.01$ L/s and $u_{out} = 0.01$ L/s denote the inlet and outlet mass flow rates, respectively. Finally, the inlet composition vector takes the values $\mathbf{W} = (W_1, ..., W_{12}) = (1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ mol/L.

The experimental data have been generated using the values of $\mathbf{a}$ and $\mathbf{Ea}$ shown in eq 12, named in what follows the true $\mathbf{a}$ and $\mathbf{Ea}$.

$$\mathbf{a} = (5 \times 10^{10}, 2 \times 10^{10}, 10^{10},$$
$$5 \times 10^9, 4 \times 10^9, 2 \times 10^9),$$
$$\mathbf{Ea} = (62733, 71097, 66914, 85734, 133829, 63987) \quad (12)$$

The concentrations obtained by solving the nonlinear system of eqs 11 with the parameters in eq 12 can be seen in Table 4 and the noisy data are shown in Table 5.

2.3.1. Example 1. In this example, the information on species E_2 and E_4 has been removed from Table 2; that is, $\mathcal{M} = \{2, 4\}$ yielding Table 6.

Table 6. Experimental Concentrations of Species E_1, E_3, E_5, E_6

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn	$<10^{-8}$		$<10^{-8}$		0.2502	$<10^{-8}$

According to Brendel's method,[4,5] none of the species in $\mathcal{M}$ can be reconstructed in this case. On the other hand, a study of structural identifiability was done. If all species are observed, the model is identifiable. The study case of structural identifiability for the parameters of the model and the unknown (no observed) concentrations E_2 and E_4 reveals that the state E_4 is not structurally identifiable, meaning that it may have an infinite number of solutions or none. This is a simple illustration of how a structurally identifiable model may convert from identifiable to

nonidentifiable when not enough information is given. Extreme concentrations are sought, that is, Problems 7 and 8 are solved by using Algorithm 1, for E_2 and E_4, yielding the results in Table 7 in which the associated parameters $\underline{\boldsymbol{\theta}_2}, \overline{\boldsymbol{\theta}_2}, \underline{\boldsymbol{\theta}_4}, \overline{\boldsymbol{\theta}_4}$ have the values in eqs 13 and 14.

$$\underline{\boldsymbol{\theta}_2} = (1650.7775, 36666.3068.6794, 10.9536, 727.8457)$$
$$\overline{\boldsymbol{\theta}_2} = (299.2406, 10^{-10}, 0.0130, 453.2508) \quad (13)$$

$$\underline{\boldsymbol{\theta}_4} = (89.9359, 10^{-10}, 0.0130, 85.7068)$$
$$\overline{\boldsymbol{\theta}_4} = (10^{-10}, 33341.9526, 10.4456, 3.6259) \quad (14)$$

We observe that there exists a wide range of the concentrations of the incomplete species E_2 and E_4 compatible with the available measurements.

In order to analyze the behavior of our approach when it deals with noisy data, we have also removed species E_2 and E_4 from Table 3, yielding Table 8

Table 8. Noisy Experimental Concentrations When $\mathcal{M} = \{2, 4\}$

	E_1	E_2	E_3	E_4	E_5	E_6
y for the true θ with noise	$<10^{-8}$		$<10^{-8}$		0.2460	$<10^{-8}$

In this case, the conclusions obtained by applying Brendel's method,[4,5] are the same as before; i.e., none of the species can be reconstructed in this case. Therefore, Problems 7 and 8 are solved, giving the extreme concentrations of Table 9, with $\underline{\boldsymbol{\theta}_2}, \overline{\boldsymbol{\theta}_2}, \underline{\boldsymbol{\theta}_4}, \overline{\boldsymbol{\theta}_4}$ in Table 9 taking the values in eq 15 and 16.

$$\underline{\boldsymbol{\theta}_2} = (3.0132, \quad 3232.6794, \quad 3.2570, \quad 1.0995)$$
$$\overline{\boldsymbol{\theta}_2} = (0.7575, \quad 10^{-10}, \quad 0.0130, \quad 0.8776) \quad (15)$$

$$\underline{\boldsymbol{\theta}_4} = (0.8313, \quad 10^{-10}, \quad 0.0130, \quad 1.0463)$$
$$\overline{\boldsymbol{\theta}_4} = (0.7205, \quad 28894.7238, \quad 9.7246, \quad 20567.2083) \quad (16)$$

Table 7. Extreme Concentrations When $\mathcal{M} = \{2, 4\}$

	E_1	E_2	E_3	E_4	E_5	E_6
y for the true $\boldsymbol{\theta}$	$<10^{-8}$	**2.6815**	$<10^{-8}$	**1.5341**	0.2502	$<10^{-8}$
y for $\underline{\boldsymbol{\theta}_2}$	$<10^{-8}$	**0.0068**	$<10^{-8}$	2.8710	0.2502	$<10^{-8}$
y for $\overline{\boldsymbol{\theta}_2}$	$<10^{-8}$	**5.7497**	$<10^{-8}$	0.5509×10^{-8}	0.2502	$<10^{-8}$
y for $\underline{\boldsymbol{\theta}_4}$	$<10^{-8}$	5.7497	$<10^{-8}$	**0.5509×10^{-8}**	0.2502	$<10^{-8}$
y for $\overline{\boldsymbol{\theta}_4}$	$<10^{-8}$	0.0071	$<10^{-8}$	**2.8713**	0.2502	$<10^{-8}$

Table 9. Extreme Concentrations When Experimental Concentrations Are Noisy and $\mathcal{M} = \{2, 4\}$

	E_1	E_2	E_3	E_4	E_5	E_6
noisy **y** for the true $\boldsymbol{\theta}$	$<10^{-8}$	**2.6974**	$<10^{-8}$	**1.5431**	0.2460	$<10^{-8}$
y for $\underline{\boldsymbol{\theta}_2}$	$<10^{-8}$	**0.0052**	$<10^{-8}$	2.8743	0.2460	$<10^{-8}$
y for $\overline{\boldsymbol{\theta}_2}$	$<10^{-8}$	**5.7539**	$<10^{-8}$	0.5517×10^{-8}	0.2460	$<10^{-8}$
y for $\underline{\boldsymbol{\theta}_4}$	$<10^{-8}$	5.7539	$<10^{-8}$	**0.5517×10^{-8}**	0.2460	$<10^{-8}$
y for $\overline{\boldsymbol{\theta}_4}$	$<10^{-8}$	0.0064	$<10^{-8}$	**2.8773**	0.2460	$<10^{-8}$

2.3.2. Example 2. In this example, the concentrations of the species E_1, E_3 and E_5 have been removed from Table 2, that is to say, $\mathcal{M} = \{1, 3, 5\}$. The observed and incomplete data in this case are shown in Table 10.

Table 10. Experimental Concentrations of Species E_2, E_4, E_6

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn		2.6815		1.5341		$<10^{-8}$

Brendel's technique,[4,5] states for this example that none of the incomplete measurements can be reconstructed, but they are all structurally identifiable, meaning that with good experimental data a unique solution should be obtained for the unknown concentrations. However, it can occur that not so good experimental data are available. Although our aim is not to reconstruct the species (it will be impossible, due to the incomplete measurements available) but to find the extreme concentrations of the unknown experimental values, in this case we see how by searching for the extreme concentrations, we determine that the minimum and the maximum concentrations take the same value, which implies the reconstruction of such species. Nevertheless, because the best experimental data have not been chosen, the reconstructed values are associated with four different sets of parameters.

The extreme concentrations obtained by solving the optimization Problems 7 and 8 when $\mathcal{M} = \{1, 3, 5\}$ and their associated parameters eq 17 and the concentrations obtained by solving the nonlinear system eq 4 with $\boldsymbol{\theta}$ in eq 9, that is, the true $\boldsymbol{\theta}$, are shown in Table 11. We can see that all the species have been

Table 11. Extreme Concentrations When $\mathcal{M} = \{1, 3, 5\}$

	E_1	E_2	E_3	E_4	E_5	E_6
y for true $\boldsymbol{\theta}$, $\underline{\boldsymbol{\theta}_1}$, $\overline{\boldsymbol{\theta}_1}$, $\underline{\boldsymbol{\theta}_3}$, $\overline{\boldsymbol{\theta}_3}$, $\underline{\boldsymbol{\theta}_5}$ and $\overline{\boldsymbol{\theta}_5}$	$<10^{-8}$	2.6815	$<10^{-8}$	1.5340	**0.2502**	$<10^{-8}$

reconstructed to their correct values, while Brendel's methodology,[4,5] was unable to reconstruct them. Furthermore, in two out of three cases, more particularly where E_1 and E_3 are minimized and maximized, we have identified the true $\boldsymbol{\theta}$ in eq 9, with three decimal digits.

$$\underline{\boldsymbol{\theta}_1} = \overline{\boldsymbol{\theta}_1} = \underline{\boldsymbol{\theta}_3} = \overline{\boldsymbol{\theta}_3} = (0.0529, 0.1280, 0.0280, 0.0001)$$

$$\underline{\boldsymbol{\theta}_5} = (545.4243, 0.1279, 0.0279, 120.6754)$$

$$\overline{\boldsymbol{\theta}_5} = (10^{-10}, 0.1280, 0.0280, 10^{-10})$$

$$(17)$$

The study of our methodology with noisy data have been analyzed in this example by removing the species E_1, E_3, and E_5 from Table 3, yielding Table 12.

Following Brendel's method,[4,5] none of the species can be reconstructed. Thus, Problems 7 and 8 are solved in order to

Table 12. Noisy Experimental Concentrations When $\mathcal{M} = \{1, 3, 5\}$

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn		2.6974		1.5431		$<10^{-8}$

Table 13. Extreme Concentrations when $\mathcal{M} = \{1, 3, 5\}$ and the Data Are Corrupted

	E_1	E_2	E_3	E_4	E_5	E_6
Noisy **y** for the true $\boldsymbol{\theta}$	$<10^{-8}$	2.6974	$<10^{-8}$	1.5431	**0.2460**	$<10^{-8}$
y for $\underline{\boldsymbol{\theta}_1}$, $\overline{\boldsymbol{\theta}_1}$, $\underline{\boldsymbol{\theta}_3}$, $\overline{\boldsymbol{\theta}_3}$, $\underline{\boldsymbol{\theta}_5}$ and $\overline{\boldsymbol{\theta}_5}$	$<10^{-8}$	2.6974	$<10^{-8}$	1.5431	**0.2163**	$<10^{-8}$

find the extreme concentrations. The solutions so-obtained can be seen in Table 13, and the associated parameters are shown in eq 18.

$$\underline{\boldsymbol{\theta}_1} = \overline{\boldsymbol{\theta}_1} = \underline{\boldsymbol{\theta}_3} = \overline{\boldsymbol{\theta}_3} = \underline{\boldsymbol{\theta}_5} = (0.0529, 0.1272, 0.0240, 0.0001)$$

$$\overline{\boldsymbol{\theta}_5} = (3893.4066, 0.1272, 0.0240, 1103.9838)$$

$$(18)$$

We observe that, in spite of the noisy data, we are able to find the values of two out of three of the incomplete measurements, by searching the highest and the lowest values that these concentrations can take. In fact, the method used in refs 4 and 5 does not allow concentrations reconstruction, but our method shows that the range of possible concentrations is degenerate, meaning that, indeed, the concentration value has been reconstructed (though we do not claim uniqueness of the parameter yielding such concentration).

2.3.3. Example 3. In the third example, the species E_2 has been removed; that is, $\mathcal{M} = \{2\}$. Hence, we assume that we have the experimental concentrations of Table 14, coming from Table 2.

Table 14. Experimental Concentrations of Species E_1, E_3, E_4, E_5, E_6

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn	$<10^{-8}$		$<10^{-8}$	1.5341	0.2502	$<10^{-8}$

The stoichiometry matrix, **A**, is given by

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ -1 & -2 & -1 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Following Brendel et al.,[4,5] we build the submatrix, $\mathbf{A_{obs}}$, of measured species,

$$\mathbf{A}_{obs} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Since the rank of $\mathbf{A}_{obs}$ is equal to the number of reactions, 4, it follows that species E_2 is identifiable, but we do not know its concentration value. Also, the generating series approach[9] reveals that E_2 is structurally identifiable, meaning that E_2 has a unique solution, obtained with a unique set of parameters if the experimental data are good enough. Therefore, we solve the optimization Problems 7 and 8 in order to find the extreme concentrations that E_2 can reach.

The concentrations obtained by solving the steady state eqs 4 respectively for the parameters $\boldsymbol{\theta}$ in eq 9, that is, the true $\boldsymbol{\theta}$, for $\underline{\boldsymbol{\theta_2}}$ and $\overline{\boldsymbol{\theta_2}}$, in eq 9 are given in Table 15.

$$\underline{\boldsymbol{\theta_2}} = (9.9336, 0.1280, 0.0279, 85.29)$$
$$\overline{\boldsymbol{\theta_2}} = (532.5699, 0.1279, 0.0279, 165.0361) \quad (19)$$

Table 15. Extreme Concentrations When $\mathcal{M} = \{2\}$

	E_1	E_2	E_3	E_4	E_5	E_6
y for the true $\boldsymbol{\theta}$, $\underline{\boldsymbol{\theta_2}}$ and $\overline{\boldsymbol{\theta_2}}$	$<10^{-8}$	**2.6815**	$<10^{-8}$	1.5341	0.2502	$<10^{-8}$

In this case, looking for the extreme concentrations of the unknown species E_2, we have found a degenerate interval with one single point, whose value is 2.6815, which corresponds to the value of the concentration of the species E_2 obtained by solving eq 4 with the true $\boldsymbol{\theta}$; that is, the incomplete concentration has been reconstructed, as Brendel et al. stated.[4,5] Nevertheless, since our goal is not to reconstruct the unknown species but to find the corresponding extreme concentrations, it makes sense that the extreme parameters are different from the true $\boldsymbol{\theta}$.

Regarding the behavior of our approach with respect to the noisy data, we have removed the species E_2 from Table 3, yielding Table 16.

Table 16. Noisy Experimental Concentrations When $\mathcal{M} = \{2\}$

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn	$<10^{-8}$		$<10^{-8}$	1.5431	0.2460	$<10^{-8}$

Problems 7 and 8 have been solved in order to find the highest and the lowest value of E_2. Such extreme concentrations and the corresponding extreme parameters can be seen in Table 17 and eq 20, respectively.

$$\underline{\boldsymbol{\theta_2}} = (255.3638, 0.1300, 0.0276, 401.1641)$$
$$\overline{\boldsymbol{\theta_2}} = (0.529, 0.1300, 0.0276, 0.0001) \quad (20)$$

This example shows that the minimum and the maximum concentrations of E_2 have the same value and equal to 2.6677. This value is different from the corresponding noisy concentration. However, we are able to fully satisfy the constraints

Table 17. Extreme Concentrations When $\mathcal{M} = \{2\}$

	E_1	E_2	E_3	E_4	E_5	E_6
noisy y for the true $\boldsymbol{\theta}$	$<10^{-8}$	**2.6974**	$<10^{-8}$	1.5431	0.2460	$<10^{-8}$
y for $\underline{\boldsymbol{\theta_2}}$ and $\overline{\boldsymbol{\theta_2}}$	$<10^{-8}$	**2.6677**	$<10^{-8}$	1.5431	0.2460	$<10^{-8}$

related with the observed concentrations, E_1, E_3, ..., E_6, in spite of the corrupted data.

2.3.4. Example 4. In this last example, species E_1, E_4, and E_5 have been deleted from Table 4, that is to say, $\mathcal{M} = \{1, 4, 5\}$, yielding Table 18.

The method of Brendel et. al,[4,5] states that in this case none of the species can be reconstructed. Therefore, the extreme concentrations of the incomplete measurements are sought by solving Problems 7 and 8. They can be seen in Table 19, and the associated parameters are in eq 21−23.

$$\underline{\mathbf{a_1}} = (5 \times 10^{10}, 2 \times 10^{10}, 9.9999 \times 10^9, 4.9999 \times 10^9,$$
$$3.9999 \times 10^9, 1.9999 \times 10^9)$$
$$\underline{\mathbf{Ea_1}} = (57557.5157, 160604.9833, 66913.9964, 85733.9848,$$
$$133815.4526, 64483.7680)$$
$$\overline{\mathbf{a_1}} = (5 \times 10^{10}, 2 \times 10^{10}, 9.9999 \times 10^9, 4.9999 \times 10^9,$$
$$4 \times 10^9, 2 \times 10^9)$$
$$\overline{\mathbf{Ea_1}} = (64678.3399, 69260.9687, 66913.9999, 85733.9993,$$
$$133857.7498, 63537.4101) \quad (21)$$

$$\underline{\mathbf{a_4}} = (5 \times 10^{10}, 2 \times 10^{10}, 9.9999 \times 10^9, 4.9999 \times 10^9,$$
$$4 \times 10^9, 2 \times 10^9)$$
$$\underline{\mathbf{Ea_4}} = (64678.3398, 69260.9692, 66913.9991, 85733.9992,$$
$$133828.9655, 63537.4074)$$
$$\overline{\mathbf{a_4}} = (4.9999 \times 10^{10}, 2 \times 10^{10}, 9.9999 \times 10^9, 5 \times 10^9,$$
$$4 \times 10^9, 2 \times 10^9)$$
$$\overline{\mathbf{Ea_4}} = (51557.5138, 164577.0411, 66913.9979,$$
$$85734.0002, 133704.0664, 64483.8224) \quad (22)$$

$$\underline{\mathbf{a_5}} = (4.9999 \times 10^{10}, 2 \times 10^{10}, 9.9999 \times 10^9, 4.9999 \times 10^9,$$
$$4 \times 10^9, 1.9999 \times 10^9)$$
$$\underline{\mathbf{Ea_5}} = (57557.5154, 138863.2615, 66913.9996, 85734.0021,$$
$$133770.9126, 64483.7748)$$
$$\overline{\mathbf{a_5}} = (5 \times 10^{10}, 2 \times 10^{10}, 10 \times 10^{10}, 5 \times 10^9, 4 \times 10^9,$$
$$2 \times 10^9)$$
$$\overline{\mathbf{Ea_5}} = (64677.8560, 69261.3722, 66914.0013, 85734.0026,$$
$$133951.3483, 63537.5477) \quad (23)$$

Finally, the behavior of our methodology with regard to the noisy data has also been analyzed in this industrial example. To test it, species E_1, E_4, and E_5 have been removed from Table 5, yielding Table 20.

As happens with the noiseless data, according to Brendel's method,[4,5] the incomplete measurements cannot be reconstructed. Thus, we look for the extreme concentrations by solving Problems 7 and 8. The solutions so-obtained are in Table 21, and their extreme parameters are shown in eq 24−26.

Table 18. Experimental Concentrations When $\mathcal{M} = \{1, 4, 5\}$

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn		0.0329	0.9587			0.0081
	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
expt concn	0.0045	0.0001	0.0001	0.23×10^{-8}	0.23×10^{-8}	0.0035

Table 19. Extreme Concentrations When $\mathcal{M} = \{1, 4, 5\}$

	E_1	E_2	E_3	E_4	E_5	E_6
y for the true (**a**, **Ea**)	**0.1791**	0.0329	0.9587	**0.8172**	**0.0689**	0.0081
y for ($\underline{a_1}$, $\underline{Ea_1}$)	**0.0412**	0.0329	0.9587	0.9552	$<10^{-8}$	0.0081
y for ($\overline{a_1}$, $\overline{Ea_1}$)	**0.2867**	0.0329	0.9587	0.7096	0.1194	0.0081
y for ($\underline{a_4}$, $\underline{Ea_4}$)	0.2867	0.0329	0.9587	**0.7096**	0.1227	0.0081
y for ($\overline{a_4}$, $\overline{Ea_4}$)	0.0412	0.0329	0.9587	**0.9552**	$<10^{-8}$	0.0081
y for ($\underline{a_5}$, $\underline{Ea_5}$)	0.0412	0.0329	0.9587	0.9552	**$<10^{-8}$**	0.0081
y for ($\overline{a_5}$, $\overline{Ea_5}$)	0.2867	0.0329	0.9587	0.7547	**0.1227**	0.0081
	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
y for the true (**a**, **Ea**)	0.0045	0.0001	0.0001	0.2333×10^{-8}	0.2333×10^{-8}	0.0035
y for ($\underline{a_1}$, $\underline{Ea_1}$)	0.0045	0.0001	0.0001	0.2243×10^{-8}	0.2243×10^{-8}	0.0035
y for ($\overline{a_1}$, $\overline{Ea_1}$)	0.0045	0.0001	0.0001	0.2312×10^{-8}	0.2312×10^{-8}	0.0035
y for ($\underline{a_4}$, $\underline{Ea_4}$)	0.0045	0.0001	0.0001	0.2333×10^{-8}	0.2333×10^{-8}	0.0035
y for ($\overline{a_4}$, $\overline{Ea_4}$)	0.0045	0.0001	0.0001	0.2426×10^{-8}	0.2426×10^{-8}	0.0035
y for ($\underline{a_5}$, $\underline{Ea_5}$)	0.0045	0.0001	0.0001	0.2376×10^{-8}	0.2376×10^{-8}	0.0035
y for ($\overline{a_5}$, $\overline{Ea_5}$)	0.0045	0.0001	0.0001	0.2245×10^{-8}	0.2245×10^{-8}	0.0035

Table 20. Noisy Experimental Concentrations When $\mathcal{M} = \{1, 4, 5\}$

	E_1	E_2	E_3	E_4	E_5	E_6
expt concn		0.0331	0.9495			0.0080
	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
expt concn	0.0044	0.0001	0.0001	0.2350×10^{-8}	0.2325×10^{-8}	0.0035

Table 21. Extreme Concentrations When $\mathcal{M} = \{1, 4, 5\}$ Using Noisy Data

	E_1	E_2	E_3	E_4	E_5	E_6
Noisy y for true (**a**, **Ea**)	**0.1755**	0.0331	0.9495	**0.8221**	**0.0678**	0.0080
y for ($\underline{a_1}$, $\underline{Ea_1}$)	**0.0405**	0.0405	0.9594	0.9594	$<10^{-8}$	$<10^{-8}$
y for ($\overline{a_1}$, $\overline{Ea_1}$)	**1**	0.0331	0.9586	$<10^{-8}$	0.4793	0.0080
y for ($\underline{a_4}$, $\underline{Ea_4}$)	1	0.0419	0.9580	**$<10^{-8}$**	0.4790	$<10^{-8}$
y for ($\overline{a_4}$, $\overline{Ea_4}$)	0.0405	0.0405	0.9594	**0.9594**	0.4852×10^{-6}	0.9498×10^{-5}
y for ($\underline{a_5}$, $\underline{Ea_5}$)	0.0417	0.0417	0.9582	0.9582	**$<10^{-8}$**	$<10^{-8}$
y for ($\overline{a_5}$, $\overline{Ea_5}$)	1	0.0405	0.9594	$<10^{-8}$	**0.4797**	$<10^{-8}$
	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
Noisy y for true (**a**, **Ea**)	0.0044	0.0001	0.0001	0.2350×10^{-8}	0.2325×10^{-8}	0.0035
y for ($\underline{a_1}$, $\underline{Ea_1}$)	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$
y for ($\overline{a_1}$, $\overline{Ea_1}$)	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$
y for ($\underline{a_4}$, $\underline{Ea_4}$)	$<10^{-8}$	1.6231×10^{-4}	1.6231×10^{-4}	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$
y for ($\overline{a_4}$, $\overline{Ea_4}$)	0.9492×10^{-5}	0.1137×10^{-4}	0.1137×10^{-4}	$<10^{-8}$	$<10^{-8}$	0.6499×10^{-8}
y for ($\underline{a_5}$, $\underline{Ea_5}$)	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$
y for ($\overline{a_5}$, $\overline{Ea_5}$)	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$

$$
\begin{aligned}
\underline{\mathbf{a_1}} \quad &= \quad (4.9999 \times 10^{10}, 2 \times 10^{10}, 9.9999 \times 10^9, 5 \times 10^9, \\
&\qquad 3.9999 \times 10^9, 1.9999 \times 10^9) \\
\underline{\mathbf{Ea_1}} \quad &= \quad (58161.8648, 282914.7484, 297418.3882, \\
&\qquad 166938.5726, 321162.7724, 113545.5758) \\
\overline{\mathbf{a_1}} \quad &= \quad (5 \times 10^{10}, 1.9999 \times 10^{10}, 9.9999 \times 10^9, 5 \times 10^9, \\
&\qquad 4 \times 10^9, 1.9999 \times 10^9) \\
\overline{\mathbf{Ea_1}} \quad &= \quad (1540044.6111, 64961.9046, 66962.7447, 85782.1988, \\
&\qquad 131955.0543, 92596.3052)
\end{aligned}
\tag{24}
$$

$$
\begin{aligned}
\underline{\mathbf{a_4}} \quad &= \quad (5 \times 10^{10}, 1.9999 \times 10^{10}, 9.9999 \times 10^9, 5 \times 10^9, \\
&\qquad 4 \times 10^9, 1.9999 \times 10^9) \\
\underline{\mathbf{Ea_4}} \quad &= \quad (355687.2217, 66464.7956, 270519.1337, \\
&\qquad 101268.7020, 303981.5657, 153259.0506) \\
\overline{\mathbf{a_4}} \quad &= \quad (4.9999 \times 10^{10}, 1.9999 \times 10^{10}, 9.9999 \times 10^9, \\
&\qquad 5 \times 10^9, 4 \times 10^9, 1.9999 \times 10^9) \\
\overline{\mathbf{Ea_4}} \quad &= \quad (58160.3235, 110197.7655, 89727.8020, \\
&\qquad 94959.1006, 160155.0092, 86871.6213)
\end{aligned}
\tag{25}
$$

$$
\begin{aligned}
\underline{\mathbf{a_5}} \quad &= \quad (5 \times 10^{10}, 2 \times 10^{10}, 10 \times 10^{10}, 4.9999 \times 10^9, \\
&\qquad 3.9999 \times 10^9, 1.9999 \times 10^9) \\
\underline{\mathbf{Ea_5}} \quad &= \quad (58362.6727, 503636.6170, 193016.0150, \\
&\qquad 204589.6321, 329745.2070, 305126.6796) \\
\overline{\mathbf{a_5}} \quad &= \quad (4.9999 \times 10^{10}, 1.9999 \times 10^{10}, 9.9999 \times 10^9, \\
&\qquad 5 \times 10^9, 3.9999 \times 10^9, 1.9999 \times 10^9) \\
\overline{\mathbf{Ea_5}} \quad &= \quad (432983.2223, 66240.8717, 555394.1069, \\
&\qquad 259001.6773, 510704.6598, 417831.0859)
\end{aligned}
\tag{26}
$$

3. AN EXTENSION. THE TRANSIENT STATE

An extension of the problem considered in section 2 is detailed here. We explain first how to extend the analysis to the evolution of the concentrations (transient state). Then, some examples are given to show the benefits of our approach.

3.1. Mathematical model. In practice it may be difficult, even impossible to measure concentrations of some reactants inside a reactor, and only their initial concentrations are known. When we do not have measurements in some time instants, or in all the time instants different from the initial one, analytical techniques, such as Brendel's method,[4,5] are useful, when applicable. However, this is not our aim, and instead this section details the problem of finding the extreme concentrations and their associated parameters in the transient state, where the evolution of the concentrations is obtained by solving the Cauchy Problem, eq 1. To experimentally address the inference problem, experimental measurements, $\widehat{y_{\mathrm{in}}}$, of the species E_i, $i = 1, ..., N$ at the time instants t_n, $n = 1, ..., S$, are taken.

As in the steady state (section 2), there may exist different sets of parameters $\boldsymbol{\theta}$ which give good fits and satisfy eq 1. This difficulty is even worse when incomplete measurements in the experimental data appear, since different concentrations curves may be coherent with the poor experimental data available.

Unlike the steady state, in the transient state we assume that we always have some information about all the species, because at least the initial concentration is known. For a fixed species E_i, $i = 1, ..., N$ we can define the S-dimensional vector, $\widehat{\mathbf{y_i}}$ as the vector of experimental concentrations of the species E_i taken in all the time instants, t_n. This vector can be divided into two parts, $\widehat{\mathbf{y_i^{obs}}}$ which represents the observed experimental data, and $\widehat{\mathbf{y_i^{mis}}}$, which denotes the concentrations that are missing.

$$
\widehat{\mathbf{y_i}} = (\widehat{\mathbf{y_i^{obs}}} | \widehat{\mathbf{y_i^{mis}}})
$$

Analogously to the steady state, for each fixed $i = 1, ..., N$, the set of time instants t_n, $n = 1, ..., S$ is split into two sets O_i and $\mathcal{M}_i$. In O_i we include the time instants for which the concentrations of the species E_i are available, and $\mathcal{M}_i$ corresponds to the time instants for which the concentrations of the species E_i are incomplete.

To calculate the incomplete measurements at a given time instant the structural approach described in refs 4 and 5 is used as a preprocessing step. This is a simple and fast algorithm based on the analysis of the rank of the stoichiometric matrix, which may allow us to obtain some reaction fluxes. When applicable, this method reconstructs the incomplete measurements perfectly. However, such methodology is limited, and unfortunately it is not always the case that incomplete data can be properly inputted from the available data. Hence, one should first restore as many data as possible in this preprocessing step, and later apply our method for the remaining missing data. As in section 2, we propose not to give one parameter vector providing a good fit, and instead we give for each species E_i, $i = 1, ..., N$ and for each missing time instant t_n, $n \in \mathcal{M}_i$ the extreme concentrations in which $\widehat{y_{\mathrm{in}}^{mis}}$ can take values, as well as the vector of parameters $\underline{\boldsymbol{\theta}_{\mathrm{in}}}$ and $\overline{\boldsymbol{\theta}_{\mathrm{in}}}$ associated with them. The extremes of the interval are the minimum and the maximum values, $\underline{y_{\mathrm{in}}}$ and $\overline{y_{\mathrm{in}}}$ respectively, that the missing concentration $\widehat{y_{\mathrm{in}}^{mis}}$ can reach. Two optimization problems have to be solved in order to find out the values of $\underline{y_{\mathrm{in}}}$ and $\overline{y_{\mathrm{in}}}$ and the associated values of $\underline{\boldsymbol{\theta}_{\mathrm{in}}}$ and $\overline{\boldsymbol{\theta}_{\mathrm{in}}}$ respectively, for $i = 1, ..., N, n \in \mathcal{M}_i$. These problems are stated in eq 27 and 28.

$$
\begin{cases}
\underline{\boldsymbol{\theta}_{\mathrm{in}}} = \arg\min y_{\mathrm{in}} \\
\text{s.t.} \quad \dot{y}(t) = \mathbf{F}(t, \mathbf{y}(t), \mathbf{u}(t), \boldsymbol{\theta}), \ t \in [0, T] \\
\qquad \mathbf{y}(0) = \mathbf{y_0} \\
\qquad |y_j(t_m) - \widehat{y_{jm}^{obs}}| \le \varepsilon, \ j = 1, ..., N, \ m \in O_j \\
\qquad y_l(t) \in \mathcal{Y}_l, \ l = 1, ..., N, \ t \in [0, T] \\
\qquad \boldsymbol{\theta} \in \Theta
\end{cases}
\tag{27}
$$

$$
\begin{cases}
\overline{\boldsymbol{\theta}_{\mathrm{in}}} = \arg\max y_{\mathrm{in}} \\
\text{s.t.} \quad \dot{y}(t) = \mathbf{F}(t, \mathbf{y}(t), \mathbf{u}(t), \boldsymbol{\theta}), \ t \in [0, T] \\
\qquad \mathbf{y}(0) = \mathbf{y_0} \\
\qquad |y_j(t_m) - \widehat{y_{jm}^{obs}}| \le \varepsilon, \ j = 1, ..., N, \ m \in O_j \\
\qquad y_l(t) \in \mathcal{Y}_l, \ l = 1, ..., N, \ t \in [0, T] \\
\qquad \boldsymbol{\theta} \in \Theta
\end{cases}
\tag{28}
$$

Table 22. Concentrations with the True θ in Equation 9

time	E_1	E_2	E_3	E_4	E_5	E_6
t_1	1	1	1	1	1	1
t_2	0.0258	2.6786	0.0257	1.4573	0.2829	0.0498
t_3	0.0006	2.6815	0.0006	1.5298	0.2518	0.0024
t_4	10^{-5}	2.6815	10^{-5}	1.5338	0.2503	0.0001

To find the global optimum of Problems 27 and 28, we first use the Brendel's method,[4,5] as a preprocessing step, not only identifying which are the species that can be reconstructed, but also giving an approach to reconstruct them. Then, Algorithm 1 is applied.

3.2. Examples. Here, we analyze the behavior of our approach in the transient state by means of two examples, detailed in sections 3.2.1 (Example 5) and 3.2.2 (Example 6). In Example 5, we select the transient state of the reaction network in eq 4 from ref 2, while in Example 6, we study the transient state of a real-world instance coming from the petrochemicals company REPSOL given by the reaction network in eq 10. The different systems of ODEs have been solved by using the function IVPRK available in the IMSL Fortran Library. The parameter R_{max} of the VNS algorithm has been set equal to 5, and the parameter $\gamma = 0.5$ in section 3.2.1, while $\gamma = 5000$ in section 3.2.2. The tolerance ε in Problems 27 and 28 is fixed to 10^{-7} in section 3.2.1 and 10^{-3} in section 3.2.2. Finally, the sets Θ and $\mathcal{Y}_l$, $l \in \mathcal{M}$ have the same expression as in section 2.3.

The concentrations, measured in mol/L, of the section 3.2.1, have been obtained by solving the system of differential equations given in eq 29 in four time instants measured in minutes: $t_1 = 0$, $t_2 = 10$, $t_3 = 20$, and $t_4 = 30$ with initial value $\mathbf{y_0} = (1, 1, 1, 1, 1, 1)$ and with the parameters θ in eq 9, that is, with the true θ are given in Table 22.

Table 24. Experimental Concentrations

time	E_1	E_2	E_3	E_4	E_5	E_6
t_1	1	1	1	1	1	1
t_2						
t_3						
t_4						

Table 25. Concentrations Obtained Using the Parameters $\underline{\theta_{5.4}}$ in Equation 31

time	E_1	E_2	E_3	E_4	E_5	E_6
t_1	1	1	1	1	1	1
t_2	0.0497	0.4053	0.0497	2.7226	0.0497	0.0497
t_3	0.0024	0.4053	0.0024	2.7936	0.0024	0.0024
t_4	0.0001	0.4053	0.0001	2.7971	**0.0001**	0.0001

$$
\begin{aligned}
\dot{y}_1(t) &= -\theta_1 y_1(t)y_2(t)y_{Cat} - u_{out}y_1(t) \\
\dot{y}_2(t) &= -\theta_1 y_1(t)y_2(t)y_{Cat} - 2\theta_2 y_2^2(t)y_{Cat} - \theta_3 y_2(t) \\
&\quad - \theta_4 y_3(t)y_2(t)y_{Cat} + W_2 u_{in} - u_{out}y_2(t) \\
\dot{y}_3(t) &= \theta_1 y_1(t)y_2(t)y_{Cat} - \theta_4 y_3(t)y_2(t)y_{Cat} - u_{out}y_3(t) \\
\dot{y}_4(t) &= \theta_2 y_2^2(t)y_{Cat} - u_{out}y_4(t) \\
\dot{y}_5(t) &= \theta_3 y_2(t) - u_{out}y_5(t) \\
\dot{y}_6(t) &= \theta_4 y_3(t)y_2(t)y_{Cat} - u_{out}y_6(t),
\end{aligned}
\tag{29}
$$

where $y_i(t)$, $i = 1, ..., N$ denotes the curve of concentration of the species E_i for $t \in [0, T]$. The vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$ yields the rate constants vector, and y_{Cat}, u_{in}, u_{out}, and W_2 have the same values as in eq 4.

Table 23. Concentrations with the True a and Ea in eq 12

time	E_1	E_2	E_3	E_4	E_5	E_6
t_1	1	1	0	0	0	0
t_2	0.5386	0.2132	0.7686	0.4613	0.1536	0.0180
t_3	0.4464	0.0955	0.8848	0.5543	0.1656	0.0194
t_4	0.4092	0.0547	0.9255	0.5904	0.1673	0.0196
t_5	0.3904	0.0367	0.9435	0.6091	0.1670	0.0196
t_6	0.3795	0.0279	0.9524	0.6199	0.1659	0.0195
t_7	0.3724	0.0235	0.9570	0.6269	0.1647	0.0193
t_8	0.3673	0.0211	0.9595	0.6320	0.1634	0.0192
t_9	0.3623	0.0199	0.9609	0.6359	0.1620	0.0190
t_{10}	0.3597	0.0193	0.9616	0.6393	0.1607	0.0189
t_{11}	0.3565	0.0191	0.9621	0.6424	0.1593	0.0187

time	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
t_1	0	0	0	0	0	0
t_2	0.0180	2.2667×10^{-5}	2.2667×10^{-5}	$<10^{-8}$	$<10^{-8}$	4.4269×10^{-5}
t_3	0.0193	2.9656×10^{-5}	2.9656×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0001
t_4	0.0194	3.3013×10^{-5}	3.3013×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0002
t_5	0.0192	3.4944×10^{-5}	3.4944×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0003
t_6	0.0190	3.6210×10^{-5}	3.6210×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0004
t_7	0.0188	3.7141×10^{-5}	3.7141×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0005
t_8	0.0185	3.7897×10^{-5}	3.7897×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0006
t_9	0.0182	3.8559×10^{-5}	3.8559×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0007
t_{10}	0.0180	3.9170×10^{-5}	3.9170×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0008
t_{11}	0.0177	3.9754×10^{-5}	3.9754×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0009

The concentrations of section 3.2.2 are also measured in mol/L, and they have been obtained by solving the system (eq 30) in 11 time instants, $t_i = 10(i − 1)$, $i = 1, ..., 11$ with the data and parameters used in section 2.3 and with $\mathbf{y_0} = (1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. Such concentrations can be seen in Table 23.

$$\dot{y}_1(t) = - a_1 \exp\left(\frac{-Ea_1}{R \times Temp}\right) y_1(t) y_2(t) y_{Cat} + \frac{1}{V}(W_1 u_{in} - u_{out} y_1(t))$$

$$\dot{y}_2(t) = - a_1 \exp\left(\frac{-Ea_1}{R \times Temp}\right) y_1(t) y_2(t) y_{Cat}$$
$$- a_2 \exp\left(\frac{-Ea_2}{R \times Temp}\right) y_2(t)^2 y_{Cat}^{0.5} - a_3 \exp\left(\frac{-Ea_3}{R \times Temp}\right) y_2(t)^2 y_{Cat}$$
$$- a_4 \exp\left(\frac{-Ea_4}{R \times Temp}\right) y_2(t) y_{Cat}^{1.1} - a_5 \exp\left(\frac{-Ea_5}{R \times Temp}\right) y_2(t) y_{Cat}^{0.5}$$
$$+ \frac{1}{V}(W_2 u_{in} - u_{out} y_2(t))$$

$$\dot{y}_3(t) = a_1 \exp\left(\frac{-Ea_1}{R \times Temp}\right) y_1(t) y_2(t) y_{Cat}$$
$$+ a_2 \exp\left(\frac{-Ea_2}{R \times Temp}\right) y_2(t)^2 y_{Cat}^{0.5} + \frac{1}{V}(W_3 u_{in} - u_{out} y_3(t))$$

$$\dot{y}_4(t) = a_1 \exp\left(\frac{-Ea_1}{R \times Temp}\right) y_1(t) y_2(t) y_{Cat}$$
$$- a_6 \exp\left(\frac{-Ea_6}{R \times Temp}\right) y_4(t) y_7(t) y_{Cat}^{1.2} + \frac{1}{V}(W_4 u_{in} - u_{out} y_4(t))$$

$$\dot{y}_5(t) = 0.5 a_2 \exp\left(\frac{-Ea_2}{R \times Temp}\right) y_2(t)^2 y_{Cat}^{0.5} + \frac{1}{V}(W_5 u_{in} - u_{out} y_5(t))$$

$$\dot{y}_6(t) = a_3 \exp\left(\frac{-Ea_3}{R \times Temp}\right) y_2(t)^2 y_{Cat} + \frac{1}{V}(W_6 u_{in} - u_{out} y_6(t))$$

$$\dot{y}_7(t) = a_3 \exp\left(\frac{-Ea_3}{R \times Temp}\right) y_2(t)^2 y_{Cat}$$
$$- a_6 \exp\left(\frac{-Ea_6}{R \times Temp}\right) y_4(t) y_7(t) y_{Cat}^{1.2} + \frac{1}{V}(W_7 u_{in} - u_{out} y_7(t))$$

$$\dot{y}_8(t) = a_4 \exp\left(\frac{-Ea_4}{R \times Temp}\right) y_2(t) y_{Cat}^{1.1} + \frac{1}{V}(W_8 u_{in} - u_{out} y_8(t))$$

$$\dot{y}_8(t) = a_4 \exp\left(\frac{-Ea_4}{R \times Temp}\right) y_2(t) y_{Cat}^{1.1} + \frac{1}{V}(W_8 u_{in} - u_{out} y_8(t))$$

$$\dot{y}_9(t) = a_4 \exp\left(\frac{-Ea_4}{R \times Temp}\right) y_2(t) y_{Cat}^{1.1} + \frac{1}{V}(W_9 u_{in} - u_{out} y_9(t))$$

$$\dot{y}_{10}(t) = a_5 \exp\left(\frac{-Ea_5}{R \times Temp}\right) y_2(t) y_{Cat}^{0.5} + \frac{1}{V}(W_{10} u_{in} - u_{out} y_{10}(t))$$

$$\dot{y}_{11}(t) = a_5 \exp\left(\frac{-Ea_5}{R \times Temp}\right) y_2(t) y_{Cat}^{0.5} + \frac{1}{V}(W_{11} u_{in} - u_{out} y_{11}(t))$$

$$\dot{y}_{12}(t) = a_6 \exp\left(\frac{-Ea_6}{R \times Temp}\right) y_4(t) y_7(t) y_{Cat}^{1.2} + \frac{1}{V}(W_{12} u_{in} - u_{out} y_{12}(t))$$

$$(30)$$

3.2.1. Example 5. We now assume that concentrations are measured only at time $t_1 = 0$, that is, the information available is the one given in Table 24. Since only initial concentrations are given, it is not possible to infer the parameters or reconstruct concentrations by Brendel's method,.[4,5] However, our method is useful to determine the extreme concentrations and infer the parameters associated with them.

As illustration, we have chosen $y_{5\,4}$ as the concentration to be minimized and maximized, that is to say, we will search for the

Table 26. Concentrations Obtained Using the Parameters $\overline{\theta}_{5\,4}$ in Equation 31

time	E_1	E_2	E_3	E_4	E_5	E_6
t_1	1	1	1	1	1	1
t_2	0.0491	0.0031	0.0490	0.0497	5.7969	0.0499
t_3	0.0024	0.0031	0.0024	0.0024	5.9868	0.0024
t_4	0.0001	0.0031	0.0001	0.0001	**5.9963**	0.0001

extreme concentrations of the species E_5 at time instant t_4 and as a byproduct the associated parameters are obtained.

The concentrations obtained by solving eq 29 with the parameters in eq 31 are shown in Tables 25 and 26.

The optimal parameters associated with Problems 27 and 28 can be seen in eq 31.

$$\begin{aligned} \underline{\theta}_{54} &= (10^{-10}, \quad 10.2158, \quad 10^{-10}, \quad 10^{-10}) \\ \overline{\theta}_{54} &= (0.7385, \quad 0.0678, \quad 562.5779, \quad 0.1566) \end{aligned} \quad (31)$$

If we compare Tables 25 and 26 we observe that the concentrations of E_1, E_3, and E_6 remain unchanged, while from the poor experimental data available (see Table 24), the uncertainty in the concentrations of E_2, E_4, and E_5 are very high. For instance, E_5 may have at time t_4 a concentration of 0.0001 or 5.9963. The first case corresponds to a model with parameters $\underline{\theta}_{54}$ and the latter with $\overline{\theta}_{54}$.

A plot with the curve $y_5(t)$, $t \in [0, 30]$ obtained as solution of eq 29 with the true parameters θ in eq 9, in black, $\underline{\theta}_{54}$ in eq 31, in green, and $\overline{\theta}_{54}$ in eq 31, in pink, is shown in Figure 1. The big points indicate the concentrations y_{5n}, $n = 1, ..., 4$.

3.2.2. Example 6. For this particular example the method described in refs 4 and 5 reveals that all the species can be perfectly reconstructed in the cases where any combination of 1, 2, 3, or 4 species is deleted from Table 23, and only in 7 combinations of 4 species such methodology cannot be applied. In order to analyze the complementarity of both methods, Brendel's approach and our algorithm, we focus on an example in which refs 4 and 5 reconstruct one out of the four missing species. Our methodology will be applied to find some information about the remaining species. In fact, in this example, we assume that the concentrations of species E_1, E_4, E_5, and E_6 have been removed, that is to say, we only have the concentrations shown in Table 27.

In this case, Brendel et al.,[4,5] state that only the species E_6 can be reconstructed. Thus, our method is very useful in order to find a possible solution of the incomplete measurements. Indeed we will find the highest and the lowest values of such incomplete concentrations of the species E_1, E_4, and E_5.

Observe that after applying Brendel's methodology,[4,5] the species E_6 is not incomplete anymore, and therefore the sets O_i and $\mathcal{M}_i$, $i = 1, ..., 12$, that we consider in Problems 27 and 28 are

$$\begin{aligned} O_i &= \{1\}, \quad \mathcal{M}_i = \{2, ..., 11\}, \quad i = 1, 4, 5 \\ O_i &= \{1, ..., 11\}, \quad \mathcal{M}_i = \varnothing, \quad i = 2, 3, 6, ..., 12 \end{aligned}$$

To show the strength of our approach, as an illustration we find the extreme values of the species E_4 at the time instant t_2; that is, we want to minimize and maximize $y_{4\,2}$. The concentrations of the species E_4 obtained after solving Problems 27 and 28 are shown in Table 28. We highlight in bold the extreme concentrations of the species E_4 at the time instant t_2. The corresponding parameters can be seen in eq 32.
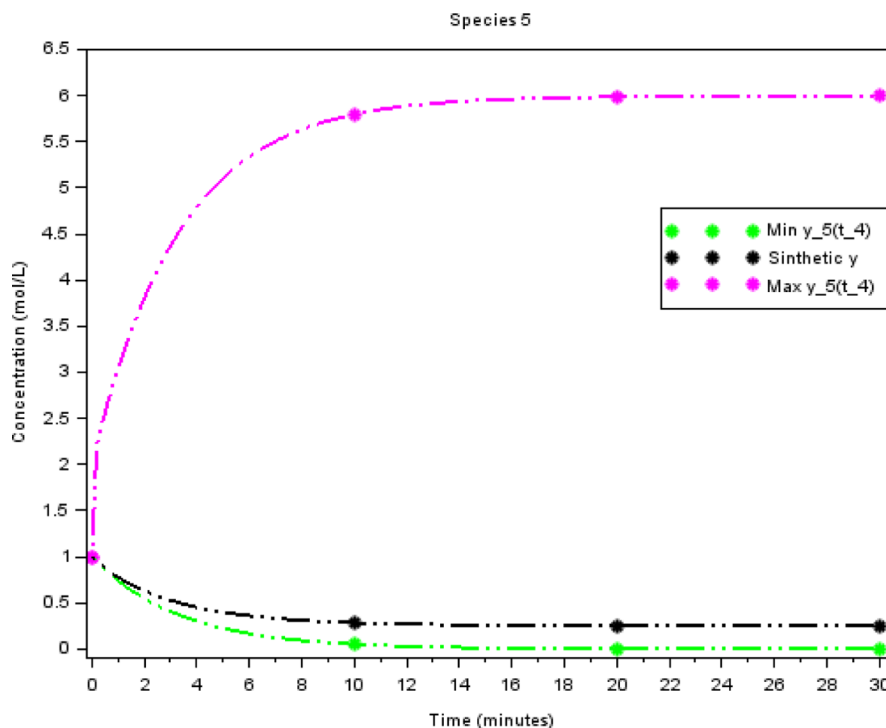
Figure 1. Concentration of species E_5 obtained with the true parameters $\boldsymbol{\theta}$ in eq 9 (black), $\underline{\boldsymbol{\theta}_{s4}}$ in eq 31 (green), and $\overline{\boldsymbol{\theta}_{s4}}$ in eq 31 (pink).

Table 27. Experimental Concentrations

time	E_1	E_2	E_3	E_4	E_5	E_6
t_1	1	1	0	0	0	0
t_2		0.2132	0.7686			
t_3		0.0955	0.8848			
t_4		0.0547	0.9255			
t_5		0.0367	0.9435			
t_6		0.0279	0.9524			
t_7		0.0235	0.9570			
t_8		0.0211	0.9595			
t_9		0.0199	0.9609			
t_{10}		0.0193	0.9616			
t_{11}		0.0191	0.9621			

time	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
t_1	0	0	0	0	0	0
t_2	0.0180	2.2667×10^{-5}	2.2667×10^{-5}	$<10^{-8}$	$<10^{-8}$	4.4269×10^{-5}
t_3	0.0193	2.9656×10^{-5}	2.9656×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0001
t_4	0.0194	3.3013×10^{-5}	3.3013×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0002
t_5	0.0192	3.4944×10^{-5}	3.4944×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0003
t_6	0.0190	3.6210×10^{-5}	3.6210×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0004
t_7	0.0188	3.7141×10^{-5}	3.7141×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0005
t_8	0.0185	3.7897×10^{-5}	3.7897×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0006
t_9	0.0182	3.8559×10^{-5}	3.8559×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0007
t_{10}	0.0180	3.9170×10^{-5}	3.9170×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0008
t_{11}	0.0177	3.9754×10^{-5}	3.9754×10^{-5}	$<10^{-8}$	$<10^{-8}$	0.0009

$$\underline{\mathbf{a}_{42}} = (4.9999 \times 10^{10}, 2 \times 10^{10}, 9.9999 \times 10^9,$$
$$5 \times 10^9, 3.9999 \times 10^9, 1.9999 \times 10^9)$$

$$\underline{\mathbf{Ea}_{42}} = (63160.8851, 70663.4134, 66922.6451,$$
$$100230.1421, 155790.0484, 64629.6122)$$

$$\overline{\mathbf{a}_{42}} = (4.9999 \times 10^{10}, 2 \times 10^{10}, 10^{10},$$
$$5 \times 10^9, 4 \times 10^9, 2 \times 10^9)$$

$$\overline{\mathbf{Ea}_{42}} = (62615.2213, 71224.4644, 66893.5414,$$
$$245633.4515, 534743.5380, 63751.6647) \tag{32}$$

In Figure 2 a plot with the curve $y_4(t), t \in [0, 100]$ obtained by solving the system (eq 30) with the true $(\mathbf{a}, \mathbf{Ea})$ in eq 12), in black, $(\underline{\mathbf{a}_{42}}, \underline{\mathbf{Ea}_{42}})$ and $(\overline{\mathbf{a}_{42}}, \overline{\mathbf{Ea}_{42}})$ in eq 32 in pink and green, respectively, is depicted. The big points denote the concentrations of y_{4n}, $n = 1, ..., 11$.

4. CONCLUSIONS

In this work we have proposed a new strategy to analyze concentrations in chemical reactions networks when experimental information is incomplete. Our scheme is not designed to provide a single set of parameters which explains the reactions network. Instead, we provide the so-called extreme concentrations,

Table 28. Concentrations of the Species E_4 Obtained with $(\underline{a_{42}}, \underline{Ea_{42}})$ and $(\overline{a_{42}}, \overline{Ea_{42}})$ in Equation 32

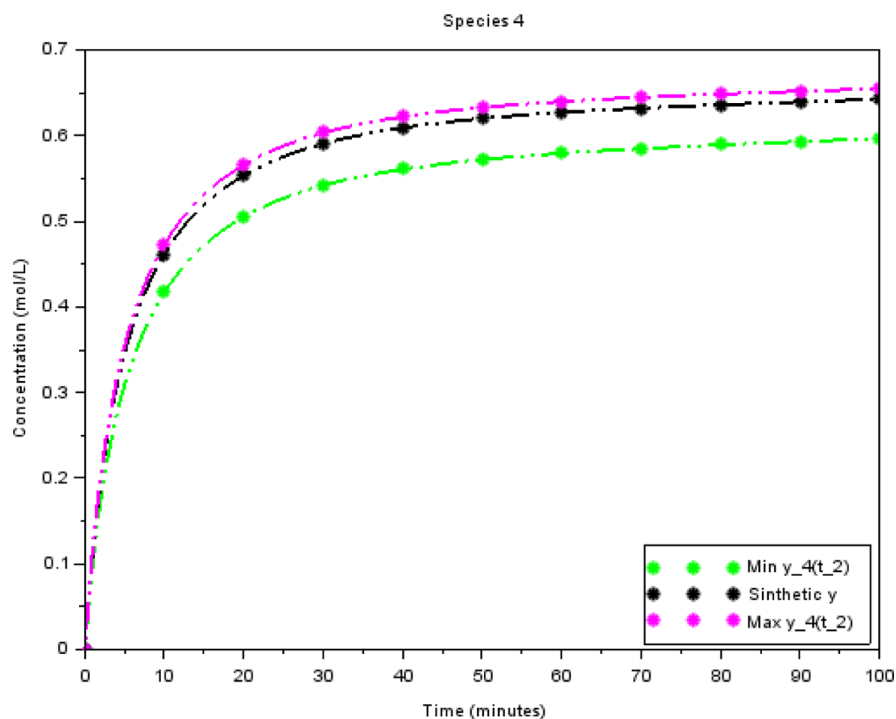|	t_1	t_2	t_3	t_4	t_5
y_4 for $(\underline{a_{42}}, \underline{Ea_{42}})$	0	**0.4175**	0.5058	0.5423	0.5611
y_4 for $(\overline{a_{42}}, \overline{Ea_{42}})$	0	**0.4736**	0.5665	0.6037	0.6223
	t_6	t_7	t_8	t_9	t_{10}	t_{11}
y_4 for $(\underline{a_{42}}, \underline{Ea_{42}})$	0.5722	0.5796	0.5850	0.5894	0.5932	0.5967
y_4 for $(\overline{a_{42}}, \overline{Ea_{42}})$	0.6330	0.6399	0.6448	0.6487	0.6520	0.6549

Figure 2. Concentration of species E_4 obtained with the true parameters $(\mathbf{a}, \mathbf{Ea})$ in eq 12 (black), $(\underline{a_{42}}, \underline{Ea_{42}})$ in eq 32 (green) and $(\overline{a_{42}}, \overline{Ea_{42}})$ in eq 32 (pink).

i.e., the lowest and highest values which can be reached by the species with such incomplete information. The resolution of a series of nonlinear optimization problems allows one to obtain not only the extreme concentrations but also the parameters associated with them. For the sake of simplicity such optimization problems are solved by using the metaheuristics VNS to avoid getting stuck at (bad) local optima. The approach is valid for both the steady-state and the transient-state cases, as illustrated in examples.

Regarding interval analysis and, in general, branch and bound methods, they have been used to address related global optimization problems.[23,24] These methods guarantee acquisition of the true global optimum of the optimization problems at the expense of high running times. Good heuristic solutions may speed up dramatically the convergence of such methods. Hence, testing how our stochastic global optimization method can be used to speed up convergence of deterministic global optimization methods is an interesting research line.

Furthermore, we have assumed throughout this paper that the kinetic laws governing the evolution of concentrations are known, though their parameters have to be inferred. However, our approach is also valid for the more general case in which uncertainty in the model does affect not only some parameters but also the kinetic law, to be taken from a given catalogue of kinetics. Designing an optimized algorithm for this challenging problem is an interesting research topic. Indeed, one has to take into account that the search of the parameters maximizing (or minimizing) the concentration of a given species must be replaced by a more complex problem of searching the kinetic law from a given catalogue and its associated parameter maximizing (or minimizing) such concentration. Such an optimization problem has now a combinatorial behavior in terms of the set of the possible kinetic laws, and thus both local and global searches must be conveniently adapted.

Moreover, we have worked under the assumption that the network is completely identified, in other words, the stoichiometry of the network is known. Usually, these stoichiometry coefficients take integer values. Nevertheless, this challenging problem in which we have to find the extreme concentrations of the incomplete measurements in the case where both integer and continuous parameters appear can be easily solved by applying our methodology, making some changes, in which the combinatorial behavior should be taken into account.

■ **AUTHOR INFORMATION**

Corresponding Author

*E-mail: asuncionjc@us.es.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Bhatt, N.; Amrhein, M.; Bonvin, D. Incremental identification of reaction and mass-transfer kinetics using the concept of extents. *Ind. Eng. Chem. Res.* **2011**, *50*, 12960−12974.

(2) Bhatt, N.; Kerimoglu, N.; Amrhein, M.; Marquardt, W.; Bonvin, D. Incremental identification of reaction systems-a comparison between rate-based and extent-based approaches. *Chem. Eng. Sci.* **2012**, *83*, 24−38.

(3) Blanquero, R.; Carrizosa, E.; Jiménez-Cordero, A.; Rodríguez, J. F. A global optimization method for model selection in chemical reactions networks. *Comput. Chem. Eng.* **2016**, *93*, 52−62.

(4) Brendel, M. *Incremental Identification of Complex Reaction Systems.* Ph.D. thesis, RWTH Aachen, 2005.

(5) Brendel, M.; Bonvin, D.; Marquardt, W. Incremental identification of kinetic models for homogeneous reaction systems. *Chem. Eng. Sci.* **2006**, *61*, 5404−5420.

(6) Burnham, S. C.; Searson, D. P; Willis, M. J; Wright, A. R Inference of chemical reaction networks. *Chem. Eng. Sci.* **2008**, *63*, 862−873.

(7) Burnham, S. C.; Willis, M. Determining reaction networks. *Comput.-Aided Chem. Eng.* **2009**, *27*, 561−566.

(8) Carrizosa, E.; Dražić, M.; Dražić, Z.; Mladenović, N. Gaussian variable neighborhood search for continuous optimization. *Comput. Oper. Res.* **2012**, *39*, 2206−2213.

(9) Chiş, O.; Banga, J. R.; Balsa-Canto, E. Genssi: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics* **2011**, *27*, 2610−2611.

(10) Chis, O.-T.; Banga, J. R.; Balsa-Canto, E. Structural identifiability of systems biology models: a critical comparison of methods. *PLoS One* **2011**, *6*, e27755.

(11) Esposito, W. R.; Floudas, C. A. Global optimization for the parameter estimation of differential-algebraic systems. *Ind. Eng. Chem. Res.* **2000**, *39*, 1291−1310.

(12) Folch-Fortuny, A.; Villaverde, A. F.; Ferrer, A.; Banga, J. R. Enabling network inference methods to handle missing data and outliers. *BMC Bioinf.* **2015**, *16*, 283.

(13) Gábor, A., Banga, J. R. Improved parameter estimation in kinetic models: Selection and tuning of regularization methods. In *Computational Methods in Systems Biology*; Springer, 2014; pp 45−60.

(14) Gábor, A.; Banga, J. R. Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Syst. Biol.* **2015**, *9*, 74.

(15) Jia, G.; Stephanopoulos, G.; Gunawan, R. Incremental parameter estimation of kinetic metabolic network models. *BMC Syst. Biol.* **2012**, *6*, 1.

(16) Laidler, K. J. *Chemical kinetics*, 3 ed.; Harper Collins, 1987.

(17) Miao, H.; Xia, X.; Perelson, A. S.; Wu, H. On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM Rev.* **2011**, *53* (1), 3−39.

(18) Mladenović, N.; Dražić, M.; Kovačevic-Vujčić, V.; Čangalović, M. General variable neighborhood search for the continuous optimization. *Eur. J. Oper. Res.* **2008**, *191*, 753−770.

(19) Mladenović, N.; Hansen, P. Variable neighborhood search. *Comput. Oper. Res.* **1997**, *24*, 1097−1100.

(20) Papamichail, I.; Adjiman, C. S. A rigorous global optimization algorithm for problems with ordinary differential equations. *J. Global. Optim.* **2002**, *24*, 1−33.

(21) Rodrigues, D.; Srinivasan, S.; Billeter, J.; Bonvin, D. Variant and invariant states for chemical reaction systems. *Comput. Chem. Eng.* **2015**, *73*, 23−33.

(22) Searson, D. P., Willis, M. J., Wright, A. Reverse engineering chemical reaction networks from time series data. *arXiv preprint arXiv:1412.6346*, 2014.

(23) Singer, A. B.; Barton, P. I. Global optimization with nonlinear ordinary differential equations. *J. Global. Optim.* **2006**, *34*, 159−190.

(24) Singer, A. B.; Taylor, J. W.; Barton, P. I.; Green, W. H. Global dynamic optimization for parameter estimation in chemical kinetics. *J. Phys. Chem. A* **2006**, *110*, 971−976.

(25) Sontag, E. D. For differential equations with r parameters, 2r+ 1 experiments are enough for identification. *J. Nonlinear Sci.* **2002**, *12*, 553−583.

(26) Swigon, D. Ensemble modeling of biological systems. *Mathematics and Life Sciences*; Walter de Gruyter, 2012; pp 19−42.

(27) Walter, E., Pronzato, L. *Identification of Parametric Models from Experimental Data*; Springer, 1997.