



Review

A global optimization method for model selection in chemical reactions networks[☆]

Rafael Blanquero^a, Emilio Carrizosa^a, Asunción Jiménez-Cordero^{a,*},
José Francisco Rodríguez^b

^a IMUS, Universidad de Sevilla, Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, C/ Tarfia s/n, 41012 Sevilla, Spain

^b Repsol Technology Center, Autovía de Extremadura s/n, 28935 Móstoles, Madrid, Spain

ARTICLE INFO

Article history:

Received 30 November 2015

Received in revised form 21 May 2016

Accepted 25 May 2016

Available online 7 June 2016

Keywords:

Model selection

Chemical reactions networks

Kinetic models

Global optimization

Variable neighborhood search

ABSTRACT

Model inference is a challenging problem in the analysis of chemical reactions networks. In order to empirically test which, out of a catalogue of proposed kinetic models, is governing a network of chemical reactions, the user can compare the empirical data obtained in one experiment against the theoretical values suggested by the models under consideration. It is thus fundamental to make an adequate choice of the decision variables (e.g. initial concentrations of the different species in the tank) in order to have maximal separation between sets of concentrations provided by the theoretical models, making then easier to identify which of the theoretical models yields data closest to those obtained empirically under identical conditions.

In this paper we illustrate how global optimization techniques can be successfully used to address the problem of model separation, as a basis for model selection. Some examples illustrate the usefulness of our approach.

© 2016 Elsevier Ltd. All rights reserved.

Contents

1. Introduction.....	52
2. Optimal initial concentrations. Problem statement.....	54
3. Initial concentrations selection. Solving the problem.....	55
3.1. Approximating the objective function.....	55
3.2. Searching an approximate ascent direction.....	55
3.3. Solving the Problem.....	56
3.4. Model selection with a large kinetic laws set.....	57
4. Some examples.....	58
4.1. Example 1.....	58
4.2. Example 2.....	59
4.3. Example 3.....	59
4.4. Example 4.....	60
5. Concluding remarks.....	61
References.....	62

[☆] Research partially supported by research grants MTM2015-65915-R (Ministerio de Ciencia e Innovación, Spain), P11-FQM-7603, FQM329 (Junta de Andalucía, Spain), all with EU ERDF funds.

* Corresponding author.

E-mail addresses: rblanquero@us.es (R. Blanquero), ecarrizosa@us.es (E. Carrizosa), asuncionjc@us.es (A. Jiménez-Cordero), jfrodriquezc@repsol.com (J.F. Rodríguez).

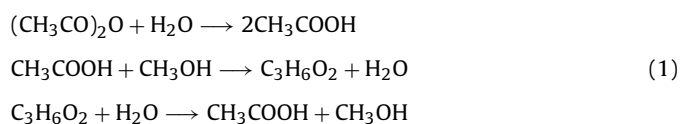
1. Introduction

In a reaction network, a set of chemical species react, making their concentrations changes along time from the initial concentrations values following a kinetic law.

Uncertainty may exist about the kinetic law itself (e.g. whether the law of mass action, Burnham and Willis (2009), Searson

et al. (2014) is governing or not the process), about its parameters (e.g. the reaction rates) or even about the reactions taking place. The user can empirically elucidate these issues by performing an experiment, introducing some initial concentrations under given conditions, letting species react, and afterwards comparing the concentrations so obtained with those which should have been obtained under identical conditions if the theoretical models were correct.

Whether model discrimination would become easier or harder may strongly depend on the choice of the experimental conditions used to compare theoretical and empirical concentrations values. To illustrate this, consider the species $S_1 = (\text{CH}_3\text{CO})_2\text{O}$ (acetic anhydride), $S_2 = \text{H}_2\text{O}$ (water), $S_3 = \text{CH}_3\text{COOH}$ (acetic acid) $S_4 = \text{CH}_3\text{OH}$ (methanol) and $S_5 = \text{C}_3\text{H}_6\text{O}_2$ (methyl acetate), which react in a batch reactor according to the following reaction network, Burnham and Willis (2009), with one reversible reaction:



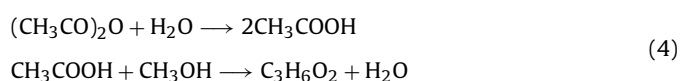
Assume that the law of mass action governs the reactions, and thus, if $x^1(t)$ is the 5-dimensional vector indicating the concentrations of species S_1, \dots, S_5 at a time instant $t \in [0, T]$, then x^1 satisfies the system of ordinary differential equations (SODE)

$$\begin{pmatrix} \dot{x}_1^1(t) \\ \dot{x}_2^1(t) \\ \dot{x}_3^1(t) \\ \dot{x}_4^1(t) \\ \dot{x}_5^1(t) \end{pmatrix} = \begin{pmatrix} -k_1 x_1^1 x_2^1 \\ -k_1 x_1^1 x_2^1 + k_2 x_3^1 x_4^1 - k_3 x_2^1 x_5^1 \\ 2k_1 x_1^1 x_2^1 - k_2 x_3^1 x_4^1 + k_3 x_2^1 x_5^1 \\ -k_2 x_3^1 x_4^1 + k_3 x_2^1 x_5^1 \\ k_2 x_3^1 x_4^1 - k_3 x_2^1 x_5^1 \end{pmatrix} \quad (2)$$

with $k_l, l = 1, 2, 3$, being the rate coefficients, taken as

$$(k_1, k_2, k_3) = (0.1, 0.15, 0.05) \text{ dm}^3 \text{ min}^{-1} \text{ mol}^{-1} \quad (3)$$

Suppose that, as an alternative model to (1), we only have the same first two reactions, i.e., the second reaction is not considered to be part of a reversible reaction:



and, as in model (1), the law of mass action is assumed, i.e., the concentrations $x^2(t)$ at time $t \in [0, T]$ under (4) is a 5-dimensional vector satisfying the SODE

$$\begin{pmatrix} \dot{x}_1^2(t) \\ \dot{x}_2^2(t) \\ \dot{x}_3^2(t) \\ \dot{x}_4^2(t) \\ \dot{x}_5^2(t) \end{pmatrix} = \begin{pmatrix} -k_1 x_1^2 x_2^2 \\ -k_1 x_1^2 x_2^2 + k_2 x_3^2 x_4^2 \\ 2k_1 x_1^2 x_2^2 - k_2 x_3^2 x_4^2 \\ -k_2 x_3^2 x_4^2 \\ k_2 x_3^2 x_4^2 \end{pmatrix} \quad (5)$$

for reaction rates (k_1, k_2) , taken here as

$$(k_1, k_2) = (0.1, 0.15) \text{ dm}^3 \text{ min}^{-1} \text{ mol}^{-1}. \quad (6)$$

Let us explore how different can be the concentrations of species S_1, \dots, S_5 along time under models (1) and (4) using identical conditions for both, as a function of the decision variables $\theta = (\theta_1, \dots, \theta_5)$ yielding the initial concentration of species $S_i, i = 1, \dots, 5$. If the vector θ of the initial concentrations takes the values

$$\theta = (0.05, 0.2, 0.11, 0.2, 0.2) \text{ dm}^{-3} \text{ mol}, \quad (7)$$

then the evolution of the concentrations of the different species S_i , and also, the difference between models (1) and (4) is absolutely clear. See Fig. 1, where the species concentrations under models (1) and (4) are depicted as solid and dashed lines respectively with the initial concentrations θ given by (7).

On the other hand, if the vector θ of initial concentrations is given by (8)

$$\theta = (0.2, 0.05, 0.2, 0.05, 0) \text{ dm}^{-3} \text{ mol}, \quad (8)$$

the difference between models (1) and (4) becomes now very weak, see Fig. 2 and compare with Fig. 1.

It would be quite easy to identify whether the empirical concentrations follow model (1) or (4), i.e., if there is or there is not a reversible reaction, in case the experiment is performed with θ as given by the former one, (7), but it would be very hard to make the discrimination if θ is chosen as the latter, (8).

The discussion above is a simple illustration of how experimental conditions (e.g. initial conditions) can help the user to discriminate between two models. In general, the problem of determining the experimental conditions which allow to discriminate among several models is known as model selection. According to Chen and Asprey (2003), there exist two points of view in the design of conventional discrimination experiments: the Bayesian methods and the frequentist methods. In Bayesian methods, Masoumi et al. (2013), Schwaab et al. (2006), Skanda and Lebiecz (2013), Box and Hill (1967), a probability is assigned to each model. This probability is updated each time an experimental design is made until there is a model whose probability is much larger than the others. One of the drawbacks of this type of methods is that usually a big number of experiments should be made in order to obtain the best model, i.e., the model with the largest probability. In practice, performing just one experimental run may be very expensive in terms of money and time, and hence we strive to obtain the optimal decision variables, e.g. the initial conditions, in one single experiment. To reach this aim the frequentist methods, Chen and Asprey (2003), Buzzi-Ferraris et al. (1990), Donckels et al. (2009), are very useful, since although they can be used in a sequential approach, good results are also obtained by making just one experiment. This is the methodology followed in this paper.

In order to measure the discrepancy among models, one has to choose one criterion. Two criteria are usually proposed in the literature, namely, the sum over all model pairs, Donckels et al. (2009), Schwaab et al. (2006), Chen and Asprey (2003), Hunter and Reiner (1965) and the worst-case method, in which a max–min approach is applied, Cooney and McDonald (1995), Bambach et al. (2013). As the motivating example of Michalik et al. (2009) shows, the first criteria may lead to a bad solution, since the proposed experiment can originate a big separation between two particular models, at the same time that some of them are very close, giving an overall good result but an inefficient model selection method. Hence, in this paper, we will use as a discrepancy measure the max–min approach. See also Donckels et al. (2010) for a multiobjective approach in which a compromise experiment is sought.

Furthermore, some papers include the possibility of a simultaneous methodology, where the parameters of the models are inferred at the same time that the best model is selected, Donckels et al. (2010), Bambach et al. (2013), Rodríguez-Fernández et al. (2013). However, we will only focus on the model-selection stage, that is to say, we assume that the models are perfectly defined by their parameters and we only want to find the experiment which best separates all the models. Parameter estimation can be made by using common strategies, see Burnham and Willis (2009), Burnham et al. (2008), Papamichail and Adjiman (2002), Singer and Barton (2006) for instance.

Moreover, although our contribution is focused on chemical reactions networks, there exist other areas modelled by dynamical

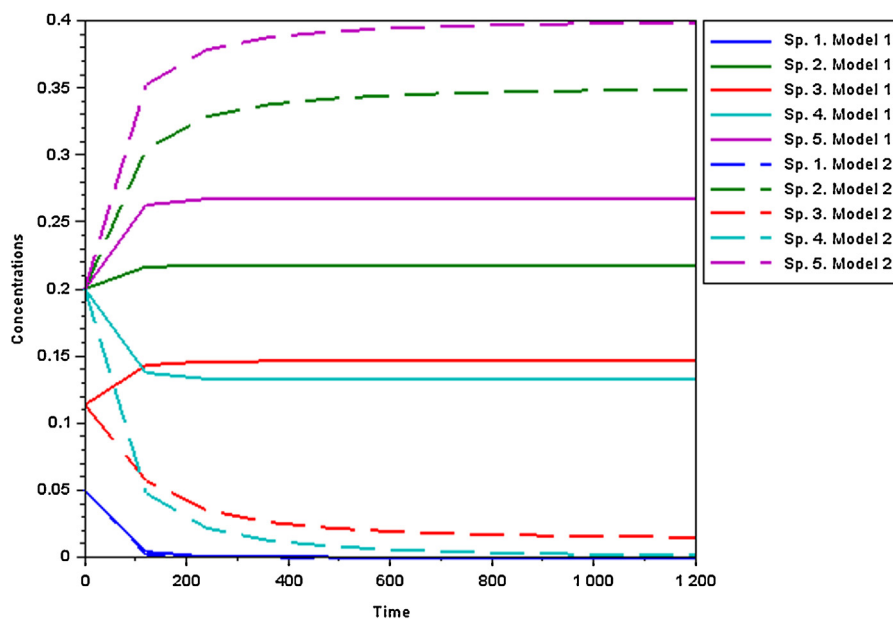


Fig. 1. Example. Models (1) and (4) with θ given in (7).

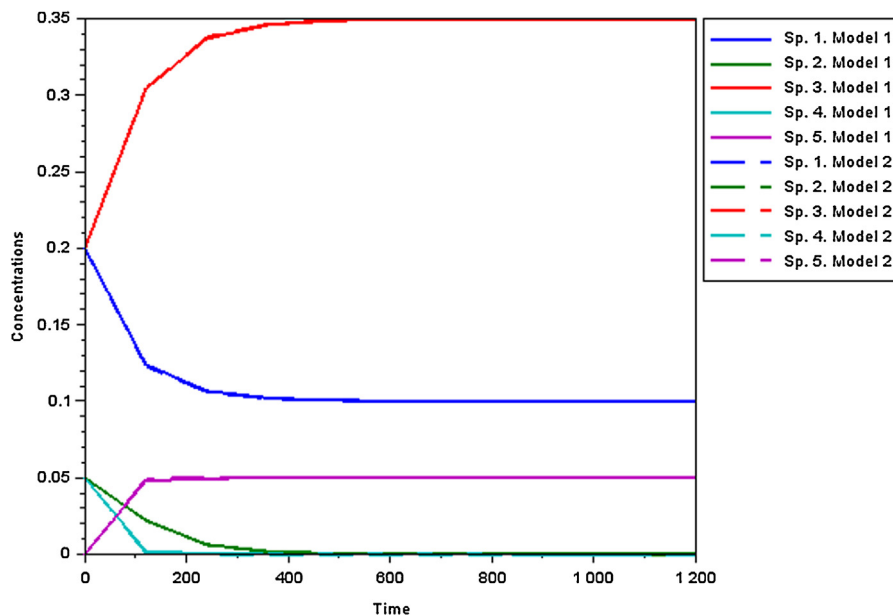


Fig. 2. Example. Models (1) and (4) with θ given in (8).

systems, where model discrimination is useful and our approach can be applied. This is the case of models which describe the epidemiology of infectious diseases. For more information, the reader is referred to [Toni et al. \(2009\)](#).

The main contribution of the paper is summarized as follows. A novel global optimization approach is proposed to choose optimal decision variables in model selection by solving a mathematical optimization problem. Our method is applicable (and more promising) when a catalogue of kinetic laws for each reaction is given, and thus an exponentially large number of models is to be separated.

The remainder of the paper is structured as follows. In Section 2, a mathematical optimization model is presented to address the problem of adequately selecting the initial concentrations yielding maximal discrimination between M models. How to solve such an optimization problem is the focus of Section 3, where a global optimization metaheuristic, namely, the variable neighborhood search,

is proposed, and some technical details on how to reduce the running times are given. Special attention is paid to detail how to address the model selection problem when a catalogue of kinetic laws is associated with each reaction, and thus a very large number of models is given. In Section 4 four examples are presented in detail, showing the usefulness of the approach. The paper finishes in Section 5 with some extensions, as well as concluding remarks.

2. Optimal initial concentrations. Problem statement

In this section we present the mathematical formulation of the problem of finding the initial concentrations of a set of species to produce the maximal discrimination between M models.

We assume that x^m , $m = 1, \dots, M$ are N -dimensional functions of the independent variable $t \in [0, T]$, which give, respectively, the

state variable under the m th model, and are assumed to be solutions of the Cauchy problems (9):

$$\begin{cases} \dot{x}^m(t, \theta) = f^m(t, x^m(t, \theta)), & t \in [0, T] \\ x^m(0, \theta) = \theta \end{cases} \quad (9)$$

Here $f^m, m=1, \dots, M$ are C^1 functions, and $\theta \in \Theta \subseteq [\underline{\theta}_1, \bar{\theta}_1] \times \dots \times [\underline{\theta}_N, \bar{\theta}_N]$ is the vector of decision variables, representing the initial conditions of the models, which is allowed to take values in the parameter space Θ , a subset of the Cartesian product of the sets of admissible values for the initial conditions θ .

Observe that, if one wants the initial concentration of species i to be fixed to some θ_i^* , one simply needs to do $\underline{\theta}_i = \bar{\theta}_i = \theta_i^*$.

For instance, for the example given in Section 1 where $M=2, f^1$ and f^2 denote respectively the right-hand sides of (2) and (5).

Given a vector $\theta \in \Theta$ of initial conditions and two indices $m_1, m_2 \in \{1, \dots, M\}$, we need to define $\Delta(x^{m_1}, x^{m_2}, \theta)$, the separation between the concentrations of the two models x^{m_1} and x^{m_2} . One sensible choice is the weighted sum of the squared ℓ_2 -distances between the curves $x_i^{m_1}$ and $x_i^{m_2}$,

$$\Delta(x^{m_1}, x^{m_2}, \theta) = \sum_{1 \leq i \leq N} \rho_i \int_0^T (x_i^{m_1}(t, \theta) - x_i^{m_2}(t, \theta))^2 dt \quad (10)$$

for some non-negative weights $\rho_i, i=1, \dots, N$, introduced to model the importance of the different species.

If $M=2$, i.e., if only two models, x^1 and x^2 are considered, our aim would be to find the values of θ maximizing $\Delta(x^1, x^2, \theta)$. When $M > 2$ models, x^1, \dots, x^M are to be separated, different criteria would be used, such as e.g. maximize the average separation, Donckels et al. (2009), Schwaab et al. (2006), Chen and Asprey (2003), Hunter and Reiner (1965), or, as suggested in this paper, maximize the minimum $\Delta(\theta, x^{m_1}, x^{m_2})$, so that all models become sufficiently separated. In other words, the optimization problem to be solved takes the form (11):

$$\begin{aligned} \max_{\theta} \psi(\theta) := & \min_{m_1, m_2 \in \{1, \dots, M\}} \Delta(x^{m_1}, x^{m_2}, \theta) \\ & m_1 \neq m_2 \\ \text{s.t. } \dot{x}^m(t, \theta) = & f^m(t, x^m(t, \theta)), \quad m = 1, \dots, M \\ x^m(0, \theta) = & \theta, \quad m = 1, \dots, M \\ \theta \in & \Theta \end{aligned} \quad (11)$$

We discuss in detail how to solve Problem (11) in Section 3.

3. Initial concentrations selection. Solving the problem

Solving Problem (11) is a challenge due to a good number of reasons. First, evaluating the objective function ψ at a given θ calls for the (numerical) resolution of the SODEs (9), and then, once $x^m(\cdot, \theta), m=1, \dots, M$ are obtained, the (numerical) evaluation of the integral defining $\Delta(x^{m_1}, x^{m_2}, \theta)$ is needed for each pair $(m_1, m_2), m_1, m_2 \in \{1, \dots, M\}, m_1 \neq m_2$. This is the focus of Section 3.1. Moreover, in order to use any local search procedure, ascent directions are to be determined. To do this, the generalized gradient of terms of the form $\int_0^T (x_i^{m_1}(t, \theta) - x_i^{m_2}(t, \theta))^2 dt$ are to be (approximately) computed, as described in Section 3.2. Since Problem (11) is likely to be multimodal, local-search procedures may get stuck at local optima. To avoid this, global optimization approaches, as the one designed in Section 3.3, are needed. Finally, the whole process becomes cumbersome, when the number M of models is large. In Section 3.4 we show how to address this case.

3.1. Approximating the objective function

The objective function ψ of Problem (11) consists of the minimum of a summation of terms of the form

$$\int_0^T (x_i^{m_1}(t, \theta) - x_i^{m_2}(t, \theta))^2 dt. \quad (12)$$

Quadrature rules allow one to approximate such integrals by weighted sums

$$\sum_{j=1}^J w_j (x_i^{m_1}(t_j, \theta) - x_i^{m_2}(t_j, \theta))^2$$

for a given grid $0 = t_1 < t_2 < \dots < t_J = T$. As a basic example of the value of w_j , we can use the composite Simpson's rule, Atkinson (1989), which states that the expression of $w_j, j=1, \dots, J$ is given by

$$w_j = \begin{cases} \frac{T}{6(J-1)}, & \text{if } j = 1, J \\ \frac{2T}{3(J-1)}, & \text{if } j \neq 1, J \text{ and } j \text{ is odd} \\ \frac{T}{3(J-1)}, & \text{if } j \neq J \text{ and } j \text{ is even} \end{cases} \quad (13)$$

Thus, we set

$$\psi(\theta) \approx \min_{\substack{m_1, m_2 \in \{1, \dots, M\} \\ m_1 \neq m_2}} \sum_{i=1}^N \sum_{j=1}^J \alpha_{ij} (x_i^{m_1}(t_j, \theta) - x_i^{m_2}(t_j, \theta))^2, \quad (14)$$

with $\alpha_{ij} = \rho_i w_j, i=1, \dots, N, j=1, \dots, J$.

3.2. Searching an approximate ascent direction

For differentiable functions, the gradient at a point gives an ascent direction (in fact, the steepest ascent one). In order to find an ascent direction for ψ , we cannot use such strategy, since the "min" operator may destroy differentiability of ψ , and the concept of generalized gradient, Clarke (1975, 1990), is used instead. In particular, given θ , let \bar{m}_1, \bar{m}_2 be the indices providing the minimum in (11), i.e., $\psi(\theta) = \Delta(x^{\bar{m}_1}, x^{\bar{m}_2}, \theta)$, an element of the generalized gradient of ψ is $\nabla \Delta(x^{\bar{m}_1}, x^{\bar{m}_2}, \theta)$. Then we use as approximate ascent direction $\nabla \Delta(x^{\bar{m}_1}, x^{\bar{m}_2}, \theta)$. In what follows we show how to compute it.

First take into account that, for $n=1, \dots, N$, and for $m_1, m_2 \in \{1, \dots, M\}$

$$\begin{aligned} \frac{\partial}{\partial \theta_n} (x_i^{m_1}(t_j, \theta) - x_i^{m_2}(t_j, \theta))^2 \\ = 2(x_i^{m_1}(t_j, \theta) - x_i^{m_2}(t_j, \theta)) \left(\frac{\partial x_i^{m_1}(t_j, \theta)}{\partial \theta_n} - \frac{\partial x_i^{m_2}(t_j, \theta)}{\partial \theta_n} \right) \end{aligned} \quad (15)$$

The next step is to find an expression of $(\partial x_i^m(t_j, \theta)) / (\partial \theta_n)$ for all n and fixed indices i, j , and m . This is not a trivial task because we recall that, for θ and m fixed, $x_i^m(\cdot, \theta)$ is the solution of a Cauchy problem which, in most of the cases, does not have a closed-form expression. The strategy followed to obtain the corresponding derivatives is

described next. We start denoting $(\partial x_i^m(t, \theta))/(\partial \theta_n)$ as $u^{imn}(t, \theta_1, \dots, \theta_n, \dots, \theta_N)$,

$$u^{imn}(t, \theta_1, \dots, \theta_n, \dots, \theta_N) = \frac{\partial x_i^m(t, \theta_1, \dots, \theta_n, \dots, \theta_N)}{\partial \theta_n},$$

$$i, n = 1, \dots, N, m = 1, \dots, M \tag{16}$$

We compute the derivative of u^{imn} with respect to t in order to build an auxiliary Cauchy problem:

$$\begin{aligned} \dot{u}^{imn}(t, \theta) &= \frac{\partial}{\partial t} \frac{\partial x_i^m}{\partial \theta_n} = \frac{\partial}{\partial \theta_n} \frac{\partial x_i^m}{\partial t} \\ &= \frac{\partial}{\partial \theta_n} f_i^m(t, x_1^m(t, \theta_1, \dots, \theta_N), \dots, x_N^m(t, \theta_1, \dots, \theta_N)) \\ &= \sum_{l=1}^N \frac{\partial f_i^m(t, x^m)}{\partial x_l^m} \frac{\partial x_l^m}{\partial \theta_n} = \sum_{l=1}^N \frac{\partial f_i^m(t, x^m)}{\partial x_l^m} u^{lmn}, \end{aligned}$$

and the initial condition of $u^{imn}(t, \theta)$ is,

$$u^{imn}(0, \theta) = \frac{\partial x_i^m(0, \theta)}{\partial \theta_n} = \frac{\partial \theta_i}{\partial \theta_n} = \delta_{in},$$

where δ_{in} denotes the Dirac delta function,

$$\delta_{in} = \begin{cases} 1, & \text{if } i = n \\ 0, & \text{otherwise} \end{cases}$$

Hence, for a fixed θ , finding the value of $u^{imn}(\cdot, \theta)$ is equivalent to solving the following Cauchy problem, for $i, n = 1, \dots, N, m = 1, \dots, M$:

$$\begin{cases} \dot{u}^{imn}(t, \theta) = \sum_{l=1}^N \frac{\partial f_i^m(t, x^m)}{\partial x_l^m} u^{lmn} \\ u^{imn}(0, \theta) = \delta_{in} \end{cases} \tag{17}$$

It is worth-mentioned that the expressions of $(\partial f_i^m)/(\partial x_l^m)$, $i, l = 1, \dots, N, m = 1, \dots, M$ are known because we have the explicit form of the functions $f_i^m(t, x^m)$, $i = 1, \dots, N, m = 1, \dots, M$.

Fixing n and θ , we can see (17) as a system of ordinary differential equations (SODE):

$$\begin{cases} \begin{pmatrix} \dot{u}^{1mn}(t, \theta) \\ \vdots \\ \dot{u}^{Nmn}(t, \theta) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1^m(t, x^m)}{\partial x_1^m} & \dots & \frac{\partial f_1^m(t, x^m)}{\partial x_N^m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N^m(t, x^m)}{\partial x_1^m} & \dots & \frac{\partial f_N^m(t, x^m)}{\partial x_N^m} \end{pmatrix} \begin{pmatrix} u^{1mn}(t, \theta) \\ \vdots \\ u^{Nmn}(t, \theta) \end{pmatrix} \\ \begin{pmatrix} \dot{u}^{1mn}(0, \theta) \\ \vdots \\ \dot{u}^{Nmn}(0, \theta) \end{pmatrix} = \mathbf{e}_n, \end{cases} \tag{18}$$

where \mathbf{e}_n is a vector with 1 in its n th coordinate and 0 everywhere else. Observe that (18) is a homogeneous linear SODE whose unknown, $u^{imn} = (u^{1mn}, \dots, u^{Nmn})$, depends on one variable, t , once the vector of the initial conditions, θ , is fixed.

Hence, for θ, m_1 and m_2 fixed, one solves the SODEs associated with $u^{im_1 n}$ and $u^{im_2 n}$ above, and then, since

$$\begin{aligned} &\frac{\partial}{\partial \theta_n} \left(\sum_{i=1}^N \sum_{j=1}^J \alpha_{ij} (x_i^{m_1}(t_j, \theta) - x_i^{m_2}(t_j, \theta))^2 \right) \\ &= \sum_{i=1}^N \sum_{j=1}^J 2\alpha_{ij} (x_i^{m_1}(t_j, \theta) - x_i^{m_2}(t_j, \theta)) (u^{im_1 n}(t_j, \theta) - u^{im_2 n}(t_j, \theta)), \end{aligned}$$

the expression (15) can be evaluated, providing an approximation to an element of the generalized gradient of $\psi(\theta)$.

3.3. Solving the Problem

We have addressed above how to (approximately) evaluate the objective function of the optimization Problem (11) as well as its generalized gradient. Hence, the optimization can be performed by means of local search procedure, e.g. Nocedal and Wright (2006), Dimitri et al. (1999). However, such local-search methods are likely to converge, at best, to a local optimum. In order to obtain the global optimum of the problem, we propose to use the well-known metaheuristic variable neighborhood search, VNS, Mladenović and Hansen (1997), adapted to optimization in continuous variables as in Mladenović et al. (2008), Carrizosa et al. (2012, 2013, 2014). The idea is to perform a local search, and, once a local optimum is obtained, to perturb in a certain neighborhood the so-obtained solution, in order to have the starting point of a new local search.

More precisely, a class of nested neighborhoods $\mathcal{N}_1(\theta) \subset \mathcal{N}_2(\theta) \subset \dots \subset \mathcal{N}_{r_{\max}}(\theta)$ of each feasible solution $\theta \in \Theta \subset \prod_{i=1}^N [\theta_i, \bar{\theta}_i]$ is built: the neighborhood $\mathcal{N}_r(\theta)$ of radius r , $1 \leq r \leq r_{\max}$, is defined as

$$\mathcal{N}_r(\theta) = \prod_{i=1}^N \left[\theta_i - \frac{r}{r_{\max}}(\theta_i - \underline{\theta}_i), \theta_i + \frac{r}{r_{\max}}(\bar{\theta}_i - \theta_i) \right] \cap \Theta,$$

This way, shaking in the largest neighborhood $\mathcal{N}_{r_{\max}}(\theta)$ amounts to sampling at random in Θ , while for small values of r , just a small perturbation of the solution is obtained.

The algorithm starts by choosing a random point $\theta \in \Theta$, and perform a local search to optimize (11) with the objective function as in (10), yielding θ^{ini} . Initially, we set $\theta^{opt} := \theta^{ini}$. Then, a local search is performed starting from some θ randomly chosen from $\mathcal{N}_1(\theta^{opt})$. This yields a new local optimum θ' . If such θ' yields better objective value than θ^{opt} , then we replace θ^{opt} by θ' and start the procedure again. Otherwise, θ' is discarded and θ^{opt} is shaken from a bigger neighborhood $\mathcal{N}_2(\theta^{opt})$. This process is repeated, by sequentially increasing the value of r , or restarted, if a better local optimum is found. A pseudocode of this process is shown in Algorithm 1.

Algorithm 1. VNS pseudocode**Algorithm 1:** VNS pseudocode**Initialization**

- Find an initial local optimal solution θ^{ini} , applying some local search method in the feasible region.
- $\theta^{opt} := \theta^{ini}$

while the stopping condition is not met **do**

$r := 1$

while $r \leq r_{max}$ **do**

 1. Define the neighborhood of radius r , as

$$\mathcal{N}_r(\theta^{opt}) = \prod_{i=1}^N \left[\theta_i^{opt} - \frac{r}{r_{max}} (\theta_i^{opt} - \underline{\theta}_i), \theta_i^{opt} + \frac{r}{r_{max}} (\overline{\theta}_i - \theta_i^{opt}) \right] \cap \Theta.$$

 2. Randomly select an initial point, θ , from the neighbourhood $\mathcal{N}_r(\theta^{opt})$.

 3. Obtain a local optimum of the problem, named θ' , starting from the solution θ .

if θ' is better than θ^{opt} **then**

$\theta^{opt} := \theta'$

 Restarting with $r := 1$

else

$r := r + 1$

end

end

end

3.4. Model selection with a large kinetic laws set

One of the steps of Algorithm 1 aims to find a local optimum of Problem (11). Observe that just the evaluation of the objective function $\psi(\theta)$ for a fixed θ implies to compute $\binom{M}{2}$ terms of the form $\Delta(x^{m_1}, x^{m_2}, \theta)$ for $m_1, m_2 \in \{1, \dots, M\}$. Evaluating $\Delta(x^{m_1}, x^{m_2}, \theta)$ amounts to solving two SODEs and evaluating numerically N integrals. If M is small, such $\binom{M}{2}$ calculations can be done in a reasonable time, but when M increases, it may be intractable to deal with such number of terms. Having a (very) large number M of models appears, in particular, when we consider a set of kinetic laws associated to each reaction of the network, as happens for instance in Brendel et al. (2006). In such cases, the number of models, M , may reach hundreds or even thousands. How to solve Problem (11) when there exists a set of kinetic laws for each reaction will be the focus of this section.

Let us consider \mathcal{C}_l , the catalogue of kinetic laws associated to the reaction l , for $l = 1, \dots, L$. The number of kinetic laws of the catalogue \mathcal{C}_l is denoted by Q_l , and each kinetic law is represented with $G_{q_l}^l$, for $q_l = 1, \dots, Q_l$. In this case, the number of models, M has the following expression:

$$M = \prod_{l=1}^L Q_l \quad (19)$$

Observe that in the catalogue \mathcal{C}_l , the no-reaction kinetic may be included, that is to say, it may exist the possibility that the reaction l does not appear at all in the reactions network, which is equivalent to saying that the reaction rate k_l is equal to zero.

In order to solve Problem (11), Algorithm 1 should be conveniently adapted to the combinatorial scheme that emerges under these conditions. Once again the VNS method is applied, but in its discrete version, Mladenović and Hansen (1997). In this context, a new neighborhood structure is to be defined. Specifically, for a combination of kinetic laws $\mathbf{G} = (q_1, \dots, q_L)$, the neighborhood of radius s , $\mathcal{M}_s(\mathbf{G})$, consists of a random selection of s reactions and an interchange between their corresponding kinetic laws and other possible laws chosen at random from their catalogues.

For a fixed θ , the algorithm begins with two initial combinations of kinetic laws, $\mathbf{G}_{m_1}^{ini}$ and $\mathbf{G}_{m_2}^{ini}$ associated to the models m_1^{ini} and m_2^{ini} , respectively, with objective value $\Delta(x^{m_1^{ini}}, x^{m_2^{ini}}, \theta)$. Furthermore, the optimal kinetic laws and the models associated are initialized as follows: $\mathbf{G}_{m_1}^{opt} := \mathbf{G}_{m_1}^{ini}$, $\mathbf{G}_{m_2}^{opt} := \mathbf{G}_{m_2}^{ini}$, $m_1^{opt} := m_1^{ini}$, and $m_2^{opt} := m_2^{ini}$. Then, the model m_1^{opt} or m_2^{opt} is randomly selected to make a perturbation of radius 1. Without loss of generality, we assume that the model selected is m_1^{opt} , and therefore the perturbation is made in $\mathcal{M}_1(\mathbf{G}_{m_1}^{opt})$. Hence, one reaction is randomly chosen and its kinetic law in $\mathbf{G}_{m_1}^{opt}$ is replaced by other kinetic law randomly selected from its catalogue, yielding a new model m'_1 . If $\Delta(x^{m'_1}, x^{m_2^{opt}}, \theta)$ is less than $\Delta(x^{m_1^{opt}}, x^{m_2^{opt}}, \theta)$, then $m_1^{opt} := m'_1$. Otherwise, a perturbation in $\mathcal{M}_2(\mathbf{G}_{m_2}^{opt})$ is made. This process is repeated until a stopping criteria, such as the maximum number of iterations, is reached. A pseudocode of the method is sketched in Algorithm 2. Observe that Algorithm 2 is included in Algorithm 1 in the search of a local optimum θ .

Algorithm 2. VNS pseudocode (discrete version).

Algorithm 2: VNS pseudocode (discrete version)

Initialization

- Fix a value of θ .
- Find an initial solution $\mathbf{G}_{m_1}^{ini}$ and $\mathbf{G}_{m_2}^{ini}$ corresponding to the models m_1^{ini} and m_2^{ini} , respectively.
- $\mathbf{G}_{m_1}^{opt} := \mathbf{G}_{m_1}^{ini}$, $\mathbf{G}_{m_2}^{opt} := \mathbf{G}_{m_2}^{ini}$, $m_1^{opt} := m_1^{ini}$ and $m_2^{opt} := m_2^{ini}$.

while the stopping condition is not met **do**

```

    s := 1
    while s ≤ smax do
        1. Choose at random the model m1opt or m2opt to be perturbed. (For the sake of simplicity, we assume in what follows that m1opt is chosen.)

        2. Obtain a perturbed model m1' from Ms(Gm1opt).

        if Δ(xm1', xm2opt, θ) is better than Δ(xm1opt, xm2opt, θ) then
            m1opt := m1', Gm1opt := Gm1'
            Restarting with s := 1
        else
            s := s + 1
        end
    end
end
end
    
```

4. Some examples

In this section, four examples are provided to show how the theoretical approach presented in the previous sections works. In all cases, the expression of Δ in (10) has $\rho_i = 1$ for all $i = 1, \dots, N$. Our 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 algorithm we have used the following values of the parameters.

- $r_{max} = 4$.
- $s_{max} = 10$

A quasi-Newton method is used as local-search routine in order to optimize the continuous variables by means of the functions `bconf`, available at the IMSL Fortran Numerical Library. Numerical integration is done as described in Section 3.1, using the weights w_j in (13).

In the examples presented in Sections 4.1–4.3, only two models are considered, i.e. $M = 2$, whereas in Section 4.4 a set of kinetic laws is associated to each reaction.

4.1. Example 1

The data come from Burnham and Willis (2009). Discrimination is to be performed between the two models given by (9), with the right hand side of the differential equation as in (2), using the values of k given in (3), and (5), where k takes the values in (6). The number of components is $N = 5$.

The interval time has the form $[0, T] = [0, 1200]$ and is measured in minutes. The number of points in the time grid is

$J = 11$, and the weights $\alpha_{ij} = \rho_i w_j = w_j$, $i = 1, \dots, N$, $j = 1, \dots, J$ are:

$$\alpha_{ij} = \begin{cases} \frac{t_{j+1} - t_j}{2}, & \text{if } j = 1 \\ \frac{t_{j+1} - t_{j-1}}{2}, & \text{if } j = 2, \dots, J - 1 \quad \forall i \\ \frac{t_j - t_{j-1}}{2}, & \text{if } j = J \end{cases} \quad (20)$$

where $t_j = 120(j - 1)$, $j = 1, \dots, J$, that is to say, the step size of the time grid is 120 min. Therefore, the weights turn out to be:

$$\alpha_{ij} = \begin{cases} 60, & \text{if } j = 1, J \\ 120, & \text{if } j = 2, \dots, J - 1 \quad \forall i \end{cases}$$

The set Θ where the initial conditions are allowed to take values is given by

$$\Theta = \{ \theta : \theta_i \in [\underline{\theta}_i, \bar{\theta}_i], i = 1, \dots, N \}, \quad (21)$$

$$\underline{\theta} = (0.05, 0.05, 0, 0.05, 0) \text{ dm}^{-3} \text{ mol},$$

$$\bar{\theta} = (0.2, 0.2, 0.2, 0.2, 0.2) \text{ dm}^{-3} \text{ mol}$$

Solving the optimization problem (11), as described in Section 3, yielded an optimal objective value of $\psi^{opt} = 68.2611$, and provided the solution θ^{opt}

$$\theta^{opt} = (0.05, 0.2, 0.1135, 0.2, 0.2) \text{ dm}^{-3} \text{ mol} \quad (22)$$

The solutions of the Cauchy problems (9) using (2), and (5) as right hand side, respectively, and taking as initial condition θ^{opt} in (22) are plotted in Fig. 1. The continuous line represents the first model, (9) with (2), while the dashed line represents the second model, (9) with (5).

As expected, in the problem of determining if the second reaction in (1) is reversible or not, the best discrimination is obtained

when the initial values of the concentrations of the species S_2 and S_5 , i.e. the reactants in the backward reaction, are high which coincides with the results in (22).

4.2. Example 2

The data for our second example are taken from Burnham et al. (2008). The two models differ in the value of the rate coefficients k^1 and k^2 , and they have the form of the Cauchy problems in (23) and (24):

$$\dot{x}^1(t, \theta) = \begin{pmatrix} -2k_1^1(x_1^1)^2 - k_2^1x_1^1 \\ k_1^1(x_1^1)^2 - k_4^1x_2^1x_4^1 \\ k_2^1x_1^1 - k_3^1x_3^1 \\ k_3^1x_3^1 - k_4^1x_2^1x_4^1 \\ k_4^1x_2^1x_4^1 \end{pmatrix}, \quad x^1(0, \theta) = \theta \quad (23)$$

and

$$\dot{x}^2(t, \theta) = \begin{pmatrix} -2k_1^2(x_1^2)^2 - k_2^2x_1^2 \\ k_1^2(x_1^2)^2 - k_4^2x_2^2x_4^2 \\ k_2^2x_1^2 - k_3^2x_3^2 \\ k_3^2x_3^2 - k_4^2x_2^2x_4^2 \\ k_4^2x_2^2x_4^2 \end{pmatrix}, \quad x^2(0, \theta) = \theta \quad (24)$$

The values of the parameters k^1 and k^2 are

$$k^1 = (0.1, 0.2, 0.13, 0.3), \quad k^2 = (0.2, 0.4, 0.26, 0.6) \quad (25)$$

The dimension of k_1^1, k_2^1, k_4^1 and k_4^2 is $\text{dm}^3 \text{min}^{-1} \text{mol}^{-1}$, while the dimension of k_3^1, k_3^2, k_3^3 and k_3^4 is min^{-1} . The number of components of the two models is $N=5$. The time interval is $[0, 15]$ minutes, which is divided in $J=16$ points. Each point of the grid, t_j , is given by

$$t_j = j - 1, \quad j = 1, \dots, J,$$

which implies that the step used is a regular step and equals to 1 min. The weights $\alpha_{ij} = \rho_i w_j = w_j$, follow the same formula as in (20), yielding

$$\alpha_{ij} = \begin{cases} 0.5, & \text{if } j = 1, J \\ 1, & \text{if } j = 2, \dots, J - 1 \end{cases} \quad \forall i$$

The expression of the set Θ is the same as in (21), but the values of $\underline{\theta}$ and $\bar{\theta}$ are:

$$\underline{\theta} = (0.33, 0.165, 0, 0, 0) \text{ dm}^{-3} \text{ mol},$$

$$\bar{\theta} = (1.34, 1.34, 1.34, 1.34, 1.34) \text{ dm}^{-3} \text{ mol}$$

Our algorithm yields an objective value of $\psi^{\text{opt}} = 7.7083$, and an optimal solution

$$\theta^{\text{opt}} = (1.34, 1.34, 1.34, 0, 0.8867) \text{ dm}^{-3} \text{ mol} \quad (26)$$

The solutions of the first model (continuous line) given by (23), and the second model (dashed line) expressed in terms of (24), can be seen in Fig. 3.

4.3. Example 3

So far, we have considered the initial concentrations as the control parameters θ . However, the very same approach, namely, modeling the discrimination problem as an optimization problem, solved by variable neighborhood search, can be used to find optimal discrimination for other types of chemical processes. To illustrate this, let us consider now the problem of model discrimination in a continuously stirred tank reactor (CSTR), in which the concentrations evolution is governed by ODEs

$$\dot{x}^m(t, \varphi, W) = f^m(t, x^m(\varphi, W)) := A\delta^m(t, x^m(\varphi, W)) + \frac{1}{V}(W\varphi - \varphi_{\text{out}}x^m(\varphi, W)) \quad (27)$$

for $m = 1, 2$, and where

$$\varphi_{\text{out}} = \sum_{p=1}^P \varphi_p,$$

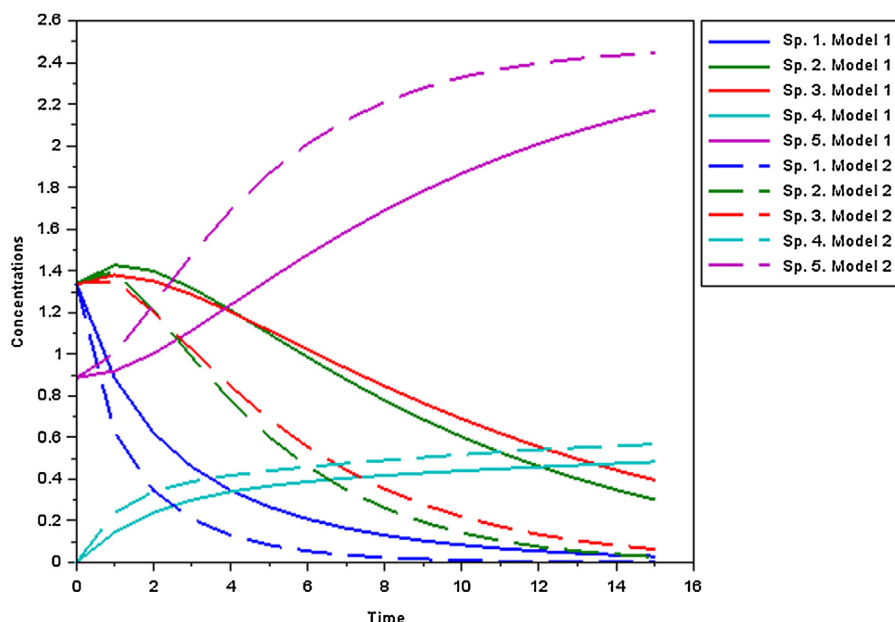


Fig. 3. Solution of Cauchy problems (23) (continued line) and (24) (dashed line) with $\theta = \theta^{\text{opt}}$ in (26).

P is the number of input streams, $A \in \mathcal{M}_{N \times L}$ is the stoichiometric matrix, N is the number of chemical species, L is the number of reactions, $V \in \mathbb{R}$ represents the volume of the tank, and δ^m , $m = 1, 2$ denote the vectors of reaction rates for x^m . Finally, the decision variable θ is the pair (φ, W) with $\varphi = (\varphi_1, \dots, \varphi_P)$ being the flow rate vector and $W \in \mathcal{M}_{N \times P}$ being the composition matrix.

As an example, we consider the following reaction network, [Bhatt et al. \(2012\)](#):



where S_i , $i = 1, \dots, 6$ are the set of species, and Z denotes a catalyst, whose concentration is $0.5 \text{ dm}^{-3} \text{ mol}$. Thus, the number of species is $N=6$ and the number of reactions is $L=4$. Furthermore, there exists one input stream ($P=1$) and the volume of the tank reactor, is $V=1 \text{ dm}^3$. The stoichiometric matrix, A , is given by

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

The expression of the reaction rates, δ^1 and δ^2 in [\(27\)](#) are given by

$$\delta^1 = \begin{pmatrix} k_1^1 x_1^1 x_2^1 x_Z^1 \\ k_2^1 (x_2^1)^2 x_Z^1 \\ k_3^1 x_2^1 \\ k_4^1 x_3^1 x_2^1 x_Z^1 \end{pmatrix}, \quad \delta^2 = \begin{pmatrix} k_1^2 x_1^2 x_2^2 x_Z^2 \\ k_2^2 (x_2^2)^2 x_Z^2 \\ k_3^2 x_2^2 \\ k_4^2 x_3^2 x_2^2 x_Z^2 \end{pmatrix}$$

where $k^1 = (0.053, 0.128, 0.028, 0.0001)$, and $k^2 = 3k^1$. The dimension of $k_1^1, k_2^1, k_4^1, k_1^2, k_2^2$ and k_4^2 is $\text{dm}^6 \text{ mol}^2 \text{ min}^{-1}$, while the dimension of k_3^1 and k_3^2 is min^{-1} . Moreover x_i^m , $i = 1, \dots, N$, $m = 1, 2$ denote the concentration of the species S_i according to the model m . Finally, the set Θ where the decision variables take values is assumed to take the form

$$\Theta = \{(\varphi, W) : \varphi_p \in [0.1, 0.5], p = 1, \dots, P; W_{ip} \in [0, 10], i = 1, \dots, N, p = 1, \dots, P\}
 \tag{29}$$

Let us formulate the optimization problem of optimal discrimination for the two models when the vector of control parameters is $\theta = (\varphi, W)$ and discrimination is to be made in the steady state; thus the following equation must be fulfilled

$$f^m(x^m(\varphi, W)) = 0, \quad m = 1, 2$$

Hence, the optimization problem to be solved is

$$\begin{aligned}
 \max_{(\varphi, W)} & \sum_{i=1}^N (x_i^1(\varphi, W) - x_i^2(\varphi, W))^2 \\
 \text{s.t.} & f^m(x^m(\varphi, W)) = 0, \quad m = 1, 2 \\
 & (\varphi, W) \in \Theta \subseteq \mathbb{R}^P \times \mathcal{M}_{N \times P}
 \end{aligned}
 \tag{30}$$

Table 1
Steady state of models [\(27\)](#) with the parameters [\(31\)](#).

Model 1	Model 2
8.258	7.236
3.979	2.401
1.741	2.761
8.418	8.605
9.931	10.111
3.383	3.384

With all these data, the optimization Problem [\(30\)](#) has been solved by using the VNS algorithm (see Section [3.3](#)). The optimal objective value is 4.642, and the optimal solution is:

$$\begin{aligned}
 \varphi^{opt} &= 0.5 \text{ dm}^3 \text{ min}^{-1}, \\
 W^{opt} &= (10, 10, 0, 6.390, 9.708, 3.382) \text{ dm}^{-3} \text{ mol}
 \end{aligned}
 \tag{31}$$

In [Table 1](#) we show the steady state of the concentrations in $\text{dm}^{-3} \text{ mol}$ of the species, S_1, \dots, S_6 , obtained after solving the non-linear systems [\(27\)](#) with the values of φ^{opt} and W^{opt} in [\(31\)](#) yielding maximal separation.

4.4. Example 4

This example considers the ODEs of the example in Section [4.3](#), which appears in [\(27\)](#) for $m = 1, \dots, M$, with $M > 2$, with the initial condition of the corresponding Cauchy problem, [\(32\)](#), as the decision variable θ and where the values of $\varphi = 0.3 \text{ dm}^3 \text{ min}^{-1}$, $W = (0, 6, 0, 0, 0, 0) \text{ dm}^{-3} \text{ mol}$, and $T = 30 \text{ min}$.

$$\begin{cases} \dot{x}^m(t, \theta) = A \delta^m(t, x^m(t, \theta)) + \frac{1}{V} (W \varphi - \varphi_{out} x^m(t, \theta)), \\ t \in [0, T], \quad m = 1, \dots, M \\ x^m(0, \theta) = \theta, \quad m = 1, \dots, M \end{cases}
 \tag{32}$$

The interval time $[0, 30]$ is discretized in $J = 151$ points, i.e. each point of the grid $t_j = 0.2(j - 1)$ and hence, the step is 0.2. The weights used follow the expression in [\(20\)](#) giving for all $i = 1, \dots, 6$,

$$\alpha_{ij} = \begin{cases} 0.1, & \text{if } j = 1, J \\ 0.2, & \text{otherwise} \end{cases}$$

The bounds of the set Θ are $\underline{\theta}_i = 0$ and $\bar{\theta}_i = 10$ for $i = 1, \dots, 6$. The reaction network used is given in [\(28\)](#). The values of the remaining parameters are taken as in Section [4.3](#).

The novelty of this example is that each component of the expression of the reaction rates δ^m comes from a set of kinetic laws associated to each reaction. In our case, we use the set of kinetic laws of [Brendel et al. \(2006\)](#), which are shown in [Table 2](#). For the

Table 2
Kinetic laws set.

Reaction 1	Reaction 2	Reaction 3	Reaction 4
$S_1 + S_2 \xrightarrow{Z} S_3$	$S_2 + S_2 \xrightarrow{Z} S_4$	$S_2 \longrightarrow S_5$	$S_3 + S_2 \xrightarrow{Z} S_6$
$G_1^1 = k_1$	$G_1^2 = k_2$	$G_3^3 = k_3$	$G_4^4 = k_4$
$G_2^1 = k_1 x_2$	$G_2^2 = k_2 x_2$	$G_2^3 = k_3 x_2$	$G_4^4 = k_4 x_2$
$G_3^1 = k_1 x_1$	$G_3^2 = k_2 x_2^2$	$G_3^3 = k_3 x_2^2$	$G_4^4 = k_4 x_3$
$G_4^1 = k_1 x_2$	$G_4^2 = k_2 x_2 x_Z$	$G_4^3 = k_3 x_2 x_Z$	$G_4^4 = k_4 x_Z$
$G_5^1 = k_1 x_2 x_1$	$G_5^2 = k_2 x_2^2 x_Z$	$G_5^3 = k_3 x_2^2 x_Z$	$G_4^4 = k_4 x_2 x_3$
$G_6^1 = k_1 x_1 x_Z$	$G_6^2 = k_2 x_Z$	$G_6^3 = k_3 x_Z$	$G_4^4 = k_4 x_3 x_Z$
$G_7^1 = k_1 x_2 x_Z$			$G_4^4 = k_4 x_2 x_Z$
$G_8^1 = k_1 x_1 x_2 x_Z$			$G_4^4 = k_4 x_1 x_3 x_Z$
$G_9^1 = k_1 x_2^2 x_Z$			$G_4^4 = k_4 x_1 x_3^2$
$G_{10}^1 = k_1 x_1 x_2^2$			$G_{10}^4 = k_4 x_2^2 x_3$

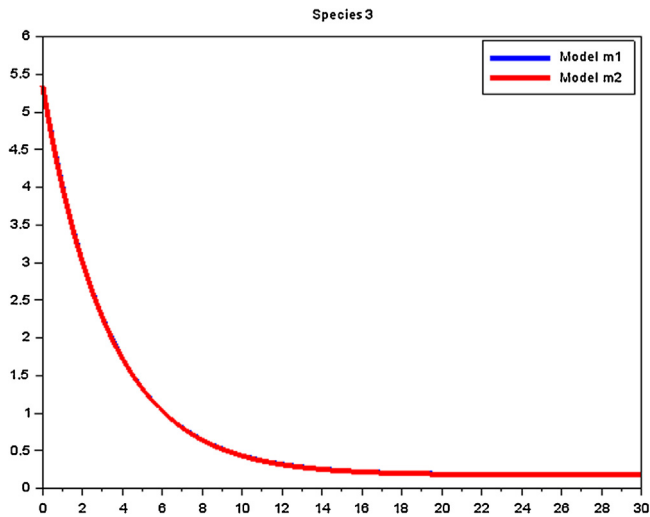


Fig. 4. Concentrations of the species S_3 with the models m_1 and m_2 given by the combination of kinetic laws in (34) and the initial concentrations in (35).

sake of simplicity, the superindex m is omitted. The values of the rate constants are

$$k = (0.053, 0.128, 0.028, 0.0001). \quad (33)$$

Therefore, from (19) we conclude that the number of possible models is $M = 10 \cdot 6 \cdot 6 \cdot 10 = 3600$. Let us show an example of δ^m . If $G_m = (3, 5, 6, 2)$, for some $m \in \{1, \dots, 3600\}$, then

$$\delta^m = \begin{pmatrix} k_1 x_1^m \\ k_2 (x_2^m)^2 x_2^m \\ k_3 x_2^m \\ k_4 x_2^m \end{pmatrix}$$

Problem (11) has been solved using Algorithms 1 and 2. The optimal value is 1.86×10^{-4} . The optimal models are

$$G_{m_1}^{opt} = (1, 5, 6, 9), \quad G_{m_2}^{opt} = (1, 5, 6, 10) \quad (34)$$

and the optimal initial conditions

$$\theta^{opt} = (9.7221, 1.2273, 5.3541, 2.8771, 3.2835, 2.8185) \quad (35)$$

We observe from (34) that the optimal models only differ in the kinetic law of the fourth reaction. This is due to the fact that we are seeking the initial concentrations that maximize the minimum difference between M models, and this minimum difference is achieved when all except one of the reactions have the same kinetic laws in both models. Furthermore, this difference is given in reaction 4 since the values of the parameters force the models G_m^4 , $m = 1, \dots, 10$ to produce very similar results, as can be seen in the small objective value. In Fig. 4, a plot of the concentrations of the species S_3 in dm^{-3} mol obtained by solving the Cauchy problem (32) with the models m_1^{opt} and m_2^{opt} given by the combination of kinetic laws in (34) and the initial conditions in (35) is depicted. As it is expected because of the parameter values, it is very difficult to distinguish both models from the plot, and therefore the model selection will be challenging in this case. Plots of the remaining species are not shown, since the conclusions obtained are very similar.

In order to analyze the influence of the fourth reaction in model selection, we performed again the experiments but omitting such reaction, which is equivalent to setting the fourth component of k in (33) equal to zero, that is to say, $k_4 = 0$. In this case, the number

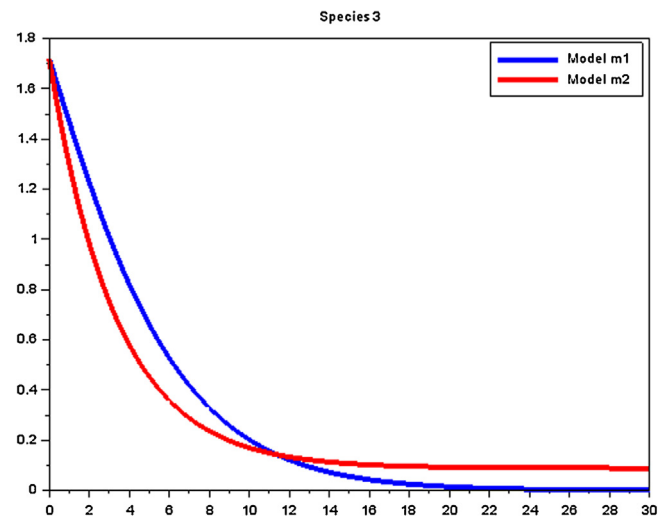


Fig. 5. Concentrations of the species S_3 with the models m_1 and m_2 given by the combination of kinetic laws in (36) and the initial concentrations in (37).

of models $M = 360$, the objective value is 1.0010 and the optimal models are given by the following combinations of kinetic laws

$$G_{m_1}^{opt} = (6, 2, 2), \quad G_{m_2}^{opt} = (4, 2, 2) \quad (36)$$

with initial concentrations

$$\theta^{opt} = (6.4956, 10, 1.7188, 2.6581, 2.8323, 2.1284) \quad (37)$$

Once more, the minimum difference between the models is achieved in two models which have two of the three reactions with the same kinetic laws, and there is only one reaction with different kinetic laws. As an illustration, the concentrations of species S_3 in dm^{-3} mol obtained after solving (32) with the models in (36) and the initial concentrations in (37), are shown in Fig. 5. We observe that, in this case, it is easier to discriminate between the models.

5. Concluding remarks

Model selection in chemical reaction networks (determine the reactions, the kinetic law, etc.) implies the knowledge of some experimental conditions (initial concentrations, flow rate, etc.). In this paper, we have proposed a method to find the decision variables that maximize the distance between theoretical models. Global optimization techniques (VNS) are used, since multiple local solutions are likely to exist. The numerical results show the strength of our approach.

The analysis performed here can be gracefully extended to discrimination problems for other reactors or other experimental conditions. Several extensions, which deserve further study, are now discussed. First, the use of alternative discrimination criteria, such as using the ℓ_1 -norm instead of the sum of squares criterion used here, would be more suitable for problems with aberrant data or heavy measurement errors. On the other hand, for simplicity we have just focused on the model discrimination problem in one stage. Extensions to multistage processes or the problem of simultaneously estimating the parameters and separating models are both nontrivial extensions which call for a deeper analysis.

From a computational point of view, the approach presented in the paper, solves in satisfactory times small instances (e.g. about 10 s for Examples 1–3) but increases for larger number of models (about 15 min for Example 4). Developing more efficient strategies, including parallel computing, is a promising research line.

References

- Atkinson, K.E., 1989. *An Introduction to Numerical Analysis*, 2nd ed. John Wiley & Sons.
- Bambach, M., Heinkenschloss, M., Herty, M., 2013. A method for model identification and parameter estimation. *Inverse Probl.* 29 (2), 025009.
- Bhatt, N., Kerimoglu, N., Amrhein, M., Marquardt, W., Bonvin, D., 2012. Incremental identification of reaction systems. A comparison between rate-based and extent-based approaches. *Chem. Eng. Sci.* 83, 24–38.
- Box, G.E.P., Hill, W.J., 1967. Discrimination among mechanistic models. *Technometrics* 9 (1), 57–71.
- Brendel, M., Bonvin, D., Marquardt, W., 2006. Incremental identification of kinetic models for homogeneous reaction systems. *Chem. Eng. Sci.* 61 (16), 5404–5420.
- Burnham, S.C., Seanson, D.P., Willis, M.J., Wright, A.R., 2008. Inference of chemical reaction networks. *Chem. Eng. Sci.* 63 (1), 862–873.
- Burnham, S.C., Willis, M.J., 2009. Determining reaction networks. *Comput. Aided Chem. Eng.* 27 (1), 561–566.
- Buzzi-Ferraris, G., Forzatti, P., Paolo, C., 1990. An improved version of a sequential design criterion for discriminating among rival multiresponse models. *Chem. Eng. Sci.* 45 (2), 477–481.
- Carrizosa, E., Dražić, M., Dražić, Z., Mladenović, N., 2012. Gaussian variable neighborhood search for continuous optimization. *Comput. Oper. Res.* 39 (9), 2206–2213.
- Carrizosa, E., Jocković, J., Ramí rez-Cobo, P., 2014. A global optimisation approach for parameter estimation of a mixture of double pareto lognormal and lognormal distributions. *Comput. Oper. Res.* 52, 231–240.
- Carrizosa, E., Olivares-Nadal, A.V., Ramí rez-Cobo, P., 2013. Time series interpolation via global optimization of moments fitting. *Eur. J. Oper. Res.* 230 (1), 97–112.
- Chen, B.H., Asprey, S.P., 2003. On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations. *Ind. Eng. Chem. Res.* 42 (7), 1379–1390.
- Clarke, F.H., 1975. Generalized gradients and applications. *Trans. Am. Math. Soc.* 205, 247–262.
- Clarke, F.H., 1990. *Optimization and Nonsmooth Analysis*. Siam.
- Cooney, M.J., McDonald, K.A., 1995. Optimal dynamic experiments for bioreactor model discrimination. *Appl. Microbiol. Biotechnol.* 43 (5), 826–837.
- Dimitri, B.P., et al., 1999. *Nonlinear Programming*. Athena Scientific, Belmont, MA.
- Donckels, B.M.R., De Pauw, D.J.W., De Baets, B., Maertens, J., Vanrolleghem, P.A., 2009. An anticipatory approach to optimal experimental design for model discrimination. *Chemom. Intell. Lab. Syst.* 95 (1), 53–63.
- Donckels, B.M.R., De Pauw, D.J.W., Vanrolleghem, P.A., De Baets, B., 2010. An ideal point method for the design of compromise experiments to simultaneously estimate the parameters of rival mathematical models. *Chem. Eng. Sci.* 65 (5), 1705–1719.
- Hunter, W.G., Reiner, A.M., 1965. Designs for discriminating between two rival models. *Technometrics* 7 (3), 307–323.
- Masoumi, S., Duever, T.A., Reilly, P.M., 2013. Sequential Markov Chain Monte Carlo (MCMC) model discrimination. *Can. J. Chem. Eng.* 91 (5), 862–869.
- Michalik, C., Stuckert, M., Marquardt, W., 2009. Optimal experimental design for discriminating numerous model candidates: the AWDC criterion. *Ind. Eng. Chem. Res.* 49 (2), 913–919.
- Nenad, M., Dražić, M., Kovačević-Vujčić, V., Čangalović, M., 2008. General variable neighborhood search for the continuous optimization. *Eur. J. Oper. Res.* 191 (3), 753–770.
- Mladenović, N., Hansen, P., 1997. Variable neighborhood search. *Comput. Oper. Res.* 24 (11), 1097–1100.
- Nocedal, J., Wright, S., 2006. *Numerical Optimization*. Springer Science & Business Media.
- Papamichail, I., Adjiman, C.S., 2002. A rigorous global optimization algorithm for problems with ordinary differential equations. *J. Glob. Optim.* 24 (1), 1–33.
- Rodríguez-Fernández, M., Rehberg, M., Kremling, A., Banga, J.R., 2013. Simultaneous model discrimination and parameter estimation in dynamic models of cellular systems. *BMC Syst. Biol.* 7 (1), 76.
- Schwaab, M., Silva, F.M., Queipo, C.A., Barreto, A.G., Nele, M., Carlos Pinto, J., 2006. A new approach for sequential experimental design for model discrimination. *Chem. Eng. Sci.* 61 (17), 5791–5806.
- Seanson, D.P., Willis, M.J., Wright, A., 2014. Reverse Engineering Chemical Reaction Networks From Time Series Data, arXiv:1412.6346.
- Singer, A.B., Barton, P.I., 2006. Global optimization with nonlinear ordinary differential equations. *J. Glob. Optim.* 34 (2), 159–190.
- Skanda, D., Lebiedz, D., 2013. A robust optimization approach to experimental design for model discrimination of dynamical systems. *Math. Program.* 141 (1–2), 405–433.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.H., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* 6 (31), 187–202.