



# Parameter estimation in ODEs: assessing the potential of local and global solvers

M. Fernández de Dios<sup>1</sup> · Ángel M. González-Rueda<sup>1,2</sup> · Julio R. Banga<sup>3</sup> · Julio González-Díaz<sup>1,2</sup> · David R. Penas<sup>3</sup>

Received: 28 June 2024 / Revised: 11 February 2025 / Accepted: 27 March 2025

© The Author(s) 2025

## Abstract

We consider the problem of parameter estimation in dynamic systems described by ordinary differential equations. A review of the existing literature emphasizes the need for deterministic global optimization methods due to the nonconvex nature of these problems. Recent works have focused on expanding the capabilities of specialized deterministic global optimization algorithms to handle more complex problems. Despite advancements, current deterministic methods are limited to problems with a maximum of around five state and five decision variables, prompting ongoing efforts to enhance their applicability to practical problems. Our study seeks to assess the effectiveness of state-of-the-art general-purpose global and local solvers in handling realistic-sized problems efficiently, and evaluating their capabilities to cope with the nonconvex nature of the underlying estimation problems.

**Keywords** Parameter estimation · Dynamic modelling · Optimization · Mathematical programming

## 1 Introduction

Mathematical models play a crucial role in describing and analyzing real-world phenomena across engineering, natural sciences, and medicine. Here we consider the problem of parameter estimation in models described by nonlinear ordinary differential equations (ODEs) (Schittkowski 2002). That is, given a parametric dynamic model and experimental data, we seek to identify the parameter values that minimize the mis-

✉ Julio González-Díaz  
julio.gonzalez@usc.es

<sup>1</sup> Department of Statistics, Mathematical Analysis and Optimization and MODESTYA Research Group, University of Santiago de Compostela, Santiago de Compostela, Spain

<sup>2</sup> CITMAGA (Galician Center for Mathematical Research and Technology), 15782 Santiago de Compostela, Spain

<sup>3</sup> Computational Biology Lab, MBG-CSIC (Spanish National Research Council), 36143 Pontevedra, Spain

match between model predictions and data. These problems play a key role in many areas, including systems biology (Mendes and Kell 1998; Ashyraliyev et al. 2009; Banga and Balsa-Canto 2008; Chou and Voit 2009) and chemical kinetics (Esposito and Floudas 2000; Singer et al. 2005).

From the mathematical point of view, parameter estimation in ODE-based models belong to the class of dynamic optimization problems. Existing solution approaches can be broadly categorized into direct and indirect methods (Biegler 2010). Indirect methods concentrate on deriving a solution that adheres to the classical necessary conditions for optimality, manifested as a two-point boundary value problem (Bryson and Ho 1969). Direct methods apply discretization schemes to transform the original infinite-dimensional problem into a finite-dimensional nonlinear programming problem. Within direct methods, sequential approaches proceed by discretizing the control variables only, transforming the problem into an outer nonlinear programming (NLP) problem with an inner initial value problem embedded (Goh and Teo 1988; Vassiliadis et al. 1994). Simultaneous approaches (Biegler 2007) discretize both control and state profiles, transforming the ODE system into algebraic equations for optimization. While the simultaneous approach offers advantages with respect to sequential approaches (e.g. easier handling of unstable dynamic systems and path constraints, and facilitating automatic differentiation), it also introduces challenges by potentially leading to large scale nonlinear programming problems that are difficult to solve.

Even though a number of local optimization methods have been developed to estimate parameters in dynamic models (Schittkowski 2002; Edsberg and Wedin 1995), they typically converge to local solutions due to the nonconvex nature of these problems (Esposito and Floudas 2000; Chen et al. 2010; Ljung and Chen 2013). A common strategy to surmount this drawback is to use a multi-start strategy with local methods (Raue et al. 2013; Fröhlich et al. 2017), i.e., to repeat local optimizations from random initial points inside the parameter's bounds. Other studies have advocated for the use of global optimization approaches, i.e., methods with specialized numerical strategies seeking the globally optimal solution (Esposito and Floudas 2000; Moles et al. 2003).

During the last two decades, several stochastic and hybrid global optimization methods have been presented, especially in the area of systems biology (Moles et al. 2003; Balsa-Canto et al. 2008; Jia et al. 2012; Egea et al. 2010; Sun et al. 2011). Even though some of these methods have shown good performance even with large-scale problems (Villaverde et al. 2015; Penas et al. 2017; Villaverde et al. 2018), an intrinsic limitation persists: their inability to guarantee global optimality. Consequently, when these techniques do not yield a satisfactory fit, it remains uncertain whether this discrepancy arises from the model's inadequacy to explain the data or from the method's inability to locate the global solution.

Given that the problem of parameter estimation in dynamic systems, and more generally, the problem of dynamic optimization, constitutes a nonconvex optimization problem, certifying global optimality requires deterministic global optimization methods. Pioneering research in this area commenced in the early 2000s (Esposito and Floudas 2000; Papamichail and Adjiman 2002, 2004; Singer et al. 2005; Chachuat et al. 2006; Lin and Stadtherr 2006, 2007; Pérez-Galván and Bogle 2017). In the area of systems biology, examples of relevant works include Polisetty et al. (2006); Panning et al. (2008); Miró et al. (2012); Pitt et al. (2018). Recent and comprehensive

overviews of the methods and challenges involved in rigorously solving these dynamic optimization problems are given by Wilhelm et al. (2019) and Song and Khan (2022).

However, as noted by Song and Khan (2022), despite all these efforts, specialized deterministic methods for global dynamic optimization are currently limited to handling problems with a maximum of approximately five state variables and five decision variables. Consequently, there is an ongoing pursuit to broaden the applicability of deterministic global optimization to dynamic systems, making them suitable for solving problems of practical significance. For example, (Sass et al. 2024) have recently presented a spatial branch-and-bound algorithm that exploits the structure of these problems when large datasets are considered.

Given the aforementioned limitations, the main goal of this paper is to study the potential of mathematical programming modeling and state-of-the-art solvers to handle a set of challenging parameter estimation problems in ODEs. Specifically, we focus on assessing the effectiveness and reliability of contemporary global and local optimization solvers. Our thorough computational experiments allow to assess to what extent global solvers can handle realistic-sized problems within acceptable computation times and, also, if local solvers can effectively tackle the nonconvex characteristics of the underlying parameter estimation problems without getting stuck at local optima.

Importantly, our numerical results show that state-of-the-art mathematical programming solvers can be effective for not-so-small problems in relatively short time (executions are limited to ten minutes): not only we find that both the local and global solvers are capable of consistently finding the optimal solutions, but we also show that the global solvers can certify global optimality. It is worth noting that these results were obtained with standard discretization schemes and out-of-the-box configurations of the solvers which, moreover, were not allowed to use any kind of parallelization capabilities. This hints at the potential of tackling even larger problems by relying on finely tuned solver configurations and high-performance computing. Last, but not least, we believe our study can serve as a benchmark to assess the potential of present and future specialized algorithms.

The rest of the paper is structured as follows. In Sect. 2 we present a detailed description of the framework for the analysis. In Sect. 3 we describe the specifics of the numerical study. Section 4 is devoted to a general overview of the computational results and, then, we devote Sect. 5 to discuss the implications of the results for relevant aspects in parameter estimation such as local optimality, identifiability and flatness of the objective function. Finally, we conclude in Sect. 6.

## 2 Framework for the analysis

In this work we focus on parameter estimation in dynamic models that can be described by a nonlinear system of differential algebraic equations, DAEs. The system depends on some parameters that we want to estimate. To this aim, some experimental data from the process is usually available (observations of the dynamic states over time), so the goal is to obtain the value of the parameters that delivers the best possible fit of the experimental data. A general approach to describe the dynamics of the model is

the following:

$$\begin{aligned}\frac{d\mathbf{x}(t)}{dt} &= f(t, \mathbf{x}(t), \mathbf{p}), \quad t \in [t_0, t_f] \\ \mathbf{y}(t) &= g(\mathbf{x}(t), \mathbf{p}), \quad t \in [t_0, t_f] \\ \mathbf{x}(t_0) &= \bar{\mathbf{x}}_0,\end{aligned}$$

where  $\mathbf{x} \in \mathbb{R}^{n_s}$  is the vector of state variables,  $\bar{\mathbf{x}}_0$  represents their initial conditions and  $\mathbf{p} \in \mathbb{R}^{n_p}$  is the vector of unknown parameters. Furthermore,  $f$  is a nonlinear function describing the dynamics of the problem and  $g$  is the observation function that gives the vector of observed states  $\mathbf{y} \in \mathbb{R}^{n_y}$  predicted by the model.

Let  $(\tau_1, \bar{\mathbf{y}}_1), \dots, (\tau_n, \bar{\mathbf{y}}_n)$  be the input experimental data, where  $\bar{\mathbf{y}}_i$  is the measured value of variable  $\mathbf{y}$  at some time  $\tau_i \in [t_0, t_f]$ . Thus, we want to solve the optimization problem of finding the value of the  $\mathbf{p}$  parameters that minimizes the error between the model predictions  $\mathbf{y}$  and the observable measurements  $\bar{\mathbf{y}}$ . This leads to the formulation of the following Non Linear Programming (NLP) problem with DAEs<sup>1</sup>:

$$\begin{aligned}\min \quad & \sum_{i=1}^n \|\mathbf{y}(\tau_i) - \bar{\mathbf{y}}_i\|^2 \\ \text{s.t.} \quad & \frac{d\mathbf{x}(t)}{dt} = f(t, \mathbf{x}(t), \mathbf{p}), \quad t \in [t_0, t_f] \\ & \mathbf{y}(t) = g(\mathbf{x}(t), \mathbf{p}), \quad t \in [t_0, t_f] \\ & \mathbf{x}(t_0) = \bar{\mathbf{x}}_0, \\ & \underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}},\end{aligned} \quad (1)$$

where  $\underline{\mathbf{p}}$  and  $\bar{\mathbf{p}}$  represent known lower and upper bounds for the parameters.

## 2.1 Direct transcription approach: **Baseline** formulation

In order to solve this problem, a direct transcription approach is employed (see Betts (2020) for details), which consists of discretizing problem (1). More precisely, the time domain is discretized into  $M = \frac{t_f - t_0}{h}$  time steps, being  $h$  the step size and  $M$  the size of the mesh. Thus, a finite dimensional approximation of the original problem is created by discretizing all the variables and constraints at discrete time points  $t_m$  with  $m \in \{0, \dots, M\}$ , where  $t_M = t_f$ . Denote by

$$\boldsymbol{\xi} = (\mathbf{x}_0, \mathbf{y}_0, \mathbf{x}_1, \mathbf{y}_1, \dots, \mathbf{x}_M, \mathbf{y}_M, \mathbf{p})$$

the optimization variables arising from the discretization, where  $\mathbf{x}_m = \mathbf{x}(t_m)$  and  $\mathbf{y}_m = \mathbf{y}(t_m)$  for all  $m \in \{0, \dots, M\}$ . Further, the ODEs are approximated using a differential numerical discretization scheme that effectively reformulates them into a

<sup>1</sup> Different error functions can be used to measure this error. We follow the standard in the field: the method of least squares.

system of algebraic equations. Therefore, the original dynamic problem (1) can be reformulated as the following standard NLP problem:

$$\begin{aligned}
 & \text{Baseline formulation} \\
 \min & \sum_{i=1}^n \|y_{m(i)} - \bar{y}_i\|^2 \\
 \text{s.t. } & H(\xi) = \mathbf{0} \\
 & \mathbf{y}_m = g(\mathbf{x}_m, \mathbf{p}), \forall m \in \{0, \dots, M\} \\
 & \mathbf{x}_0 = \bar{\mathbf{x}}_0 \\
 & \underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}},
 \end{aligned} \tag{2}$$

where  $H$  is the system of algebraic equations obtained after applying the chosen discretization scheme and  $m(i)$  gives the correspondence between the times at which the observable measures are taken,  $\tau_i$ , and the discretization points  $t_m$ , i.e., for each  $i \in \{1 \dots, n\}$ ,  $m(i)$  gives the value such that  $t_{m(i)} = \tau_i$ .<sup>2</sup> Naturally, the larger  $M$  is, the better (2) approximates problem (1). For our numerical study we implemented five classic discretization schemes (see Burden and Faires (2001)). For the sake of completeness, we present below the resulting approximation of the ODEs:

- The Euler method. For each  $m \in \{0, \dots, M - 1\}$  we have

$$\mathbf{x}_{m+1} = \mathbf{x}_m + hf(t_m, \mathbf{x}_m, \mathbf{p}). \tag{3}$$

- The Trapezoid method. For each  $m \in \{0, \dots, M - 1\}$  we have

$$\mathbf{x}_{m+1} = \mathbf{x}_m + \frac{h}{2}(f(t_m, \mathbf{x}_m, \mathbf{p}) + f(t_{m+1}, \mathbf{x}_{m+1}, \mathbf{p})). \tag{4}$$

- The Adams-Moulton method with step 3. For each  $m \in \{0, \dots, M - 3\}$  we have<sup>3</sup>

$$\begin{aligned}
 \mathbf{x}_{m+3} = \mathbf{x}_{m+2} + \frac{h}{24} & \left( 9f(t_{m+3}, \mathbf{x}_{m+3}, \mathbf{p}) + 19f(t_{m+2}, \mathbf{x}_{m+2}, \mathbf{p}) \right. \\
 & \left. - 5f(t_{m+1}, \mathbf{x}_{m+1}, \mathbf{p}) + f(t_m, \mathbf{x}_m, \mathbf{p}) \right).
 \end{aligned} \tag{5}$$

- The Simpson method. For each  $m \in \{1, \dots, M - 1\}$  we have<sup>4</sup>

$$\mathbf{x}_{m+1} = \mathbf{x}_{m-1} + \frac{h}{3}(f(t_{m+1}, \mathbf{x}_{m+1}, \mathbf{p}) + 4f(t_m, \mathbf{x}_m, \mathbf{p}) + f(t_{m-1}, \mathbf{x}_{m-1}, \mathbf{p})). \tag{6}$$

<sup>2</sup> Thus, we are implicitly assuming that the discrete time points of the observable measures  $\tau_i$  are a subset of the discretization points  $t_m$ . If this were not the case, an interpolation approximation could be applied to recover this property.

<sup>3</sup> Initial conditions for  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are taken from the Trapezoid method.

<sup>4</sup> Initial conditions for  $\mathbf{x}_1$  are taken from the Trapezoid method.

- The Runge-Kutta method. For each  $m \in \{1, \dots, M - 1\}$  we have

$$\begin{aligned}
 \mathbf{x}_{m+1} &= \mathbf{x}_m + \frac{h}{6}(\mathbf{k}_m^1 + 2\mathbf{k}_m^2 + 2\mathbf{k}_m^3 + \mathbf{k}_m^4), \text{ where} \\
 \mathbf{k}_m^1 &= f(t_m, \mathbf{x}_m, \mathbf{p}) \\
 \mathbf{k}_m^2 &= f(t_m + \frac{h}{2}, \mathbf{x}_m + \frac{h}{2} \cdot \mathbf{k}_1, \mathbf{p}) \\
 \mathbf{k}_m^3 &= f(t_m + \frac{h}{2}, \mathbf{x}_m + \frac{h}{2} \cdot \mathbf{k}_2, \mathbf{p}) \\
 \mathbf{k}_m^4 &= f(t_{m+1}, \mathbf{x}_m + h \cdot \mathbf{k}_3, \mathbf{p}).
 \end{aligned} \tag{7}$$

## 2.2 Variations of the mathematical programming model

Given the difficulty of the optimization problems to be solved, we also consider two standard variations of the `Baseline` formulation, with the goal of assessing the impact of these alternative formulations in the performance of the state-of-the-art solvers.

First, we define a relaxation of `Baseline` by introducing a fixed feasibility tolerance in the algebraic constraints. Specifically, given  $\epsilon > 0$ , the `ExtraTol` formulation is given by

$$\begin{aligned}
 &\text{ExtraTol formulation} \\
 \min &\sum_{i=1}^n \|\mathbf{y}_{m(i)} - \bar{\mathbf{y}}_i\|^2 \\
 \text{s.t.} &-\epsilon \leq H(\boldsymbol{\xi}) \leq \epsilon \\
 &\mathbf{y}_m = g(\mathbf{x}_m, \mathbf{p}), \forall m \in \{0, \dots, M\} \\
 &\mathbf{x}_0 = \bar{\mathbf{x}}_0 \\
 &\underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}}.
 \end{aligned} \tag{8}$$

Note that feasibility tolerances can usually be controlled with solver-specific options. Yet, `ExtraTol` is solver independent, which enables a direct comparison of the impact of these tolerances across solvers without having to deal with the particularities of their respective configuration options.

A second variation of `Baseline` formulation is to treat all the algebraic constraints of the model as soft constraints. This is achieved by adding, to each algebraic constraint, a couple of slack variables, which are then heavily penalized in the objective function. In this case, we obtain `SoftCons` formulation given by

$$\begin{aligned}
 &\text{SoftCons formulation} \\
 \min &\sum_{i=1}^n \|\mathbf{y}_{m(i)} - \bar{\mathbf{y}}_i\|^2 + P \cdot \sum_{j=1}^{n_1} s_j
 \end{aligned}$$

$$\begin{aligned}
& \text{s.t. } -\mathbf{s} \leq H(\boldsymbol{\xi}) \leq \mathbf{s} \\
& \mathbf{y}_m = g(\mathbf{x}_m, \mathbf{p}), \quad \forall m \in \{0, \dots, M\} \\
& \mathbf{x}_0 = \bar{\mathbf{x}}_0 \\
& \underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}} \\
& \mathbf{s} \geq \mathbf{0},
\end{aligned} \tag{9}$$

where  $\mathbf{s} \in \mathbb{R}^{n_1}$  are the slack variables added to the algebraic equations of the ODEs system. Further,  $P \geq 0$  is the penalization in the objective function for using the slack variables, i.e., the penalization imposed on the violations of the constraints.

### 3 Setting up the numerical study

The proposed framework has been implemented in AMPL (Fourer et al. 1990), using a PERL script (Wall et al. 2000) to automatically generate, for each problem instance, the AMPL model files for the different NLP formulations and the different discretization schemes. All numerical experiments have been performed on the supercomputer Finisterrae III, provided by Galicia Supercomputing Centre (CESGA), using computational nodes powered with two 32-core Intel Xeon Ice Lake 8352Y CPUs with 256GB of RAM and 1TB SSD.<sup>5</sup>

#### 3.1 Pool of parameter estimation problems

We have applied our framework to a series of well known parameter estimation problems from the literature, which we outline below (refer to Appendix A for a deeper description):

- `alpha pinene`: it represents the mechanism for thermal isomerization of  $\alpha$ -pinene (Fuguitt and Hawkins 1945). It has 5 parameters and 5 states (fully observed).
- `BBG` (Biomass batch growth): it describes the microbial growth in a stirred fed-batch bioreactor (Rodriguez-Fernandez et al. 2007). It has 4 parameters and 2 states (fully observed).
- `FHN`: the Fitzhugh-Nagumo model (FitzHugh 1961; Nagumo et al. 1962). It has 3 parameters and 2 states with only one observed state (partially observed).
- `harmonic`: it represents a harmonic oscillator (pendulum), as considered in Go et al. (2023). It has 2 parameters and 2 states (fully observed).
- `Lotka Volterra`: it is a two species Lotka-Volterra predatory-prey model, with 3 parameters and 2 states, as considered by Go et al. (2023). Two versions of this problem are considered depending on the number of observed variables considered when solving it. We denote by `Lotka VolterraF` the case with 2 observed states (fully observed) and by `Lotka VolterraP` the problem with only one observed state (partially observed).

<sup>5</sup> It is worth noting that NEOS Server (Czyzyk et al. 1998) was also intensively used during the initial stages of the project to test and validate different model formulations.

**Table 1** Number of state variables, observed variables, parameters and time points for measurements

Problem	State Var $n_s$	Observed Var $n_y$	Parameters $n_p$	Observed times $n$
alpha pinene	5	5	5	8
BBG	2	2	4	7
FHN	2	1	3	6
harmonic	2	2	2 (2)	10
Lotka Volterra <sup>F</sup>	2	2	3 (2)	20
Lotka Volterra <sup>P</sup>	2	1	3 (2)	20
daisy mamil3 <sup>F</sup>	3	3	5 (3)	20
daisy mamil3 <sup>P</sup>	3	2	5 (3)	20
hiv <sup>F</sup>	5	5	10 (5)	20
hiv <sup>P</sup>	5	4	10 (5)	20
Crauste <sup>F</sup>	5	5	13 (5)	20
Crauste <sup>P</sup>	5	4	13 (5)	20

- `daisy mamil`: it is based on the examples used in the analysis of DAISY identifiability software (Bellu et al. 2007), and it represents a 3-compartment model. It has 5 parameters and 3 states. Two versions of this problem are considered: `daisy mamil3F` (fully observed) and `daisy mamil3P` (with partially observable data: 2 observed states).
- `hiv`: it is based on a model of HIV infection dynamics coming from (Wodarz and Nowak 1999). It has 10 parameters and 5 states. Two versions of this problem are considered: `hivF` (fully observed) and `hivP` (with partially observable data: 3 observed states and observed data for the sum of the other 2 states).
- `Crauste`: it is based on the system described in (Crauste et al. 2017), and it has 16 parameters and 5 state variables. Two versions of this problem are considered: `CrausteF` (fully observed) and `CrausteP` (with partially observable data: 3 observed states and observed data for the sum of the other 2 states).

The procedure to generate the noiseless measurement data for the observed variables of each problem is as follows.<sup>6</sup> We take nominal values for the parameters of the different systems of ODES from past literature and solve the ODE system for those values of the parameters with the R package `pracma` (Borchers 2023) (using Runge–Kutta). Then, from the results given by `pracma`, we select a sample of points of the state variables. As we describe in Appendix A, for most of the problems these points are uniformly distributed on a given time of the form  $[0, T]$ , with  $T$  ranging from 1 to 20. Table 1 shows, for each problem, the number of i) state variables, ii) observed variables, iii) parameters<sup>7</sup> and iv) sample time points.

It is worth mentioning that, in the literature, most of these problems are solved under fixed initial conditions. However, for many of our problems (such as `harmonic` and

<sup>6</sup> The only exception is `alpha pinene` which, given its simplicity, was fed with real (and noisy) measurements.

<sup>7</sup> When appropriate, we indicate in parenthesis the number of initial conditions to be estimated.

the four problems for which we have both fully and partially versions), we assume these conditions are unknown and need to be estimated as well. This increases the difficulty for solving them.

### 3.2 Configurations of the elements of the computational framework

Each parameter estimation problems was tackled with different configurations of the elements involved in our framework. We now detail all such configurations:

- I. Mathematical formulations: we run the three mathematical programming formulations described in the previous section: `Baseline`, `ExtraTol` and `SoftCons`. The configurable parameters for `ExtraTol` and `SoftCons` are chosen as follows:
  - For `ExtraTol`, two values for the feasibility tolerance  $\epsilon$  are considered:  $10^{-4}$  and  $10^{-6}$ .<sup>8</sup>
  - For `SoftCons`, two values for  $P$  are considered:  $10^3$  and  $10^5$ .
- II. Discretization numerical schemes: the 5 discretization schemes introduced in the previous section are considered: `Euler`, `Trapezoid`, `Adams–Moulton`, `Simpson` and `Runge–Kutta`. Further, each problem is run with two different mesh sizes ( $M$ ), as specified in Table 2.<sup>9</sup>
- III. Mathematical programming solvers:] given the interest in analyzing whether the solutions obtained with our framework are global optima, it is pertinent to distinguish between local and global solvers. Specifically, three global solvers and five local solvers are employed in the numerical experiments:
 

Global Solvers: `Couenne`<sup>G</sup> 0.5.7 (Belotti et al. 2009), `BARON`<sup>G</sup> 21.1.13 (2021.01.13) (Tawarmalani and Sahinidis 2005) and `Ocateract`<sup>G</sup> Engine v4.4.1 (Ocateract Optimisation Intelligence 2023).

Local Solvers: `BONMIN`<sup>L</sup> 1.8.7 (Bonami et al. 2012), `Knitro`<sup>L</sup> 13.2.0 (Byrd et al. 2006), `Ipopt`<sup>L</sup> 3.12.13 (Wächter and Biegler 2006), `CONOPT`<sup>L</sup> 3.17A (Drud 1985) and `SNOPT`<sup>L</sup> 7.5–1.2 (Gill et al. 2005).

Each solver is allowed to use only one thread and a time limit of 10 min is imposed. All other options of the solvers are set to their default values. All the combinations of the above configurations result in a total of 5280 executions. In particular, we tested at least 400 configurations on each parameter estimation problem: 5 mathematical programming formulations, 5 discretization schemes with 2 mesh sizes each and 8 state-of-the-art solvers.

### 3.3 Performance metrics

In order to evaluate the quality of the solutions given by our framework, two metrics are employed. On the one hand, it is important to check if the reference solution is

<sup>8</sup> In order to avoid some numerical issues observed in some preliminary runs, slightly different options are considered for `Lotka` `Volterra`<sup>P</sup> ( $10^{-5}$ ,  $10^{-6}$  and  $10^{-8}$ ) and `Crauste` ( $10^{-5}$ ,  $10^{-7}$  and  $10^{-9}$ ).

<sup>9</sup> For `Lotka` `Volterra`<sup>P</sup> problem, on top of the mesh sizes of 100 and 1000, a bigger mesh of size 10000 was also tested.

**Table 2** Mesh sizes for each problem

Lotka	Volterra	Crauste	FHN	harmonic	alpha pinene	BBG	daisy	mamil	hiv
100		100	200	230	1230	120	100		100
1000		1000	2000	2300	3690	1200	1000		1000

found. Given a problem, let  $\mathbf{p}^{\text{REF}}$  be the reference solution of the parameters and let  $\mathbf{p}$  be the estimation obtained with our approach. Then, the *maximum of the relative error* is computed as follows:

$$\text{MaxRE} = \max_{i=1, \dots, n_p} \left| \frac{p_i - p_i^{\text{REF}}}{p_i^{\text{REF}}} \right|.$$

Note that, the smaller MaxRE is, the closer of the solution to the reference one. In the numerical study, the criterion to consider that the solution is acceptably close to the reference solution is that  $\text{MaxRE} < 0.1$ . The error showed in the figures of the computational results corresponds to MaxRE.

On the other hand, since some of the problems can present a potential lack of identifiability, it could happen that the parameter estimation obtained is not really close to the reference solution, but the adjustment with respect to the observed state variables is good. In order to analyze this, we use the following *normalized mean squared error* measure:

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^n \|\mathbf{y}_{m(i)} - \bar{\mathbf{y}}_i\|^2}{n_y \cdot n}}}{\bar{y}_{\max} - \bar{y}_{\min}},$$

where  $\bar{y}_{\max} = \max\{\max\{\bar{\mathbf{y}}_1\}, \dots, \max\{\bar{\mathbf{y}}_n\}\}$  and  $\bar{y}_{\min} = \min\{\min\{\bar{\mathbf{y}}_1\}, \dots, \min\{\bar{\mathbf{y}}_n\}\}$ .

Recall that, with the exception of alpha pinene, all problems are fed with noiseless synthetic data, whose generation ensures that the NRMSE at the reference solution should be zero or very close to it (depending on the precision of the discretization).

## 4 Numerical results: general overview

This section focuses on the study of the overall performance of the mathematical programming solvers and the different modeling choices. The results are illustrated with a series of tables, each of them containing the following columns:

SOLVED<sup>S</sup>. Proportion of runs in which the solver returned “solved” as its final status.  
 FOUND<sup>R</sup>. Proportion of runs in which the solver actually found the reference solution ( $\text{MaxRE} \leq 0.1$ ).

NEAR<sup>R</sup>. Proportion of runs in which the solver did not find the reference solution but was close to it ( $0.1 < \text{MaxRE} \leq 0.5$ ).

ALTERN. Proportion of runs in which the solution was not close to the reference solution ( $\text{MaxRE} > 0.5$ ) but, still,  $\text{NRMSE} < 0.0001$  (potential lack of identifiability).

$\text{TIME}^{\text{BFR}}$ . Smallest solve time, in seconds, among the runs in which the solver found the reference solution.

SUCCESS. Proportion of runs in which the solver execution was successful, in the sense of not experimenting numerical issues, having abrupt terminations or failing to return a feasible solution.

A supplementary Online Appendix (<https://bit.ly/ParamEstimODEs>) contains a more comprehensive set of results, including more complete tables and additional figures.

#### 4.1 Results by problem

Table 3 shows the results disaggregated by problem and sorted by column  $\text{FOUND}^{\text{R}}$ . A first finding that stands out is that, for each and every problem, the reference solution was found at least once. Moreover, in more than half of the problems, the reference solution was found by more than one third of the configurations, which shows the robustness of the mathematical programming formulations, since no fine tuning of the underlying configurations is needed to find the reference solutions and even certifying its global optimality. Column  $\text{TIME}^{\text{BFR}}$  shows that running times are also promising since, with the exception of *alpha pinene*, for each problem there is at least one configuration that has found the reference solution in less than one second and, very often, in less than a tenth of a second. These running times are particularly encouraging, given that they include problems such as *Crauste* and *hiv*, which have more than 10 parameters to estimate. Yet, having quick running times should not come as a surprise, since many of the configurations involve relatively coarse discretizations. What is truly interesting is that, despite this coarseness, the reference solution is often found.

The results in Table 3 also hint at two other aspects particularly relevant in parameter estimation problems and that we explore more deeply in Sect. 5. First, in some problems the  $\text{SOLVED}^{\text{S}}$  value is higher than the value in  $\text{FOUND}^{\text{R}}$ , suggesting that there are multiple local optima in the problem at hand. At the same time, a value in  $\text{FOUND}^{\text{R}}$  higher than the one in  $\text{SOLVED}^{\text{S}}$  indicates that global solvers have found the optimal solution, but failed to certify it. Second, values different from zero in ALTERN come from the existence of high quality solutions different from the reference one, which may be the result of a lack of identifiability in the problem at hand; indeed, with the exception of *Crauste*<sup>F</sup>, the largest values in ALTERN are in partially specified problems.

#### 4.2 Results by solver

We move now to Table 4, which shows the results by solver, again sorted by column  $\text{FOUND}^{\text{R}}$ . The first aspect that stands out is that two of the global optimization solvers involved in the study come out on top, which is particularly valuable given that they

**Table 3** Summary results by problem

Problem	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
daisy_mamil3 <sup>F</sup>	0.735	0.885	0.022	0.002	0.029	0.965
Lotka_Volterra <sup>F</sup>	0.743	0.835	0.022	0	0.026	0.960
harmonic	0.715	0.733	0.002	0	0.032	0.927
hiv <sup>F</sup>	0.820	0.688	0.122	0.082	0.033	0.975
BBG	0.595	0.438	0.052	0	0.084	0.802
alpha_pinene	0.250	0.365	0.005	0	1.786	0.667
daisy_mamil3 <sup>P</sup>	0.665	0.343	0.092	0.030	0.028	0.958
Crauste <sup>F</sup>	0.796	0.298	0.173	0.362	0.054	0.958
hiv <sup>P</sup>	0.792	0.195	0.082	0.490	0.103	0.965
Crauste <sup>P</sup>	0.781	0.185	0.144	0.375	0.069	0.969
FHN	0.415	0.117	0.007	0.030	0.514	0.802
Lotka_Volterra <sup>P</sup>	0.694	0.024	0.376	0.393	0.033	0.867

**Table 4** Summary results by solver

Solver	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
BARON <sup>G</sup>	0.667	0.518	0.135	0.168	0.076	0.992
Couenne <sup>G</sup>	0.691	0.482	0.130	0.089	0.138	0.841
Knitro <sup>L</sup>	0.961	0.430	0.130	0.156	0.026	0.994
Ocateract <sup>G</sup>	0.524	0.408	0.115	0.182	1.286	0.867
BONMIN <sup>L</sup>	0.897	0.403	0.111	0.117	0.038	0.897
Ipopt <sup>L</sup>	0.886	0.395	0.124	0.211	0.038	0.933
CONOPT <sup>L</sup>	0	0.306	0.098	0.215	0.044	0.921
SNOPT <sup>L</sup>	0.752	0.221	0.045	0.212	0.040	0.764

are often capable of providing global optimality certificates for the discretization at hand. As expected, local solvers are faster, although BARON<sup>G</sup> is very competitive.

### 4.3 Results by discretization scheme

In Table 5 we present the results by discretization scheme and, although there are no major differences, Trapezoid seems to perform slightly better than the rest.

### 4.4 Results by mathematical programming formulation

Finally, Table 6 studies the impact of the variations of the mathematical programming models to be solved. Again, the performance seems to be similar across the three approaches and, if anything, the use of SoftCons may be slightly worse.

**Table 5** Summary results by discretization scheme

Scheme	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
Trapezoid	0.723	0.462	0.104	0.177	0.026	0.951
Adams-Moulton	0.693	0.399	0.104	0.160	0.029	0.891
Simpson	0.691	0.396	0.126	0.149	0.028	0.914
Runge-Kutta	0.604	0.365	0.097	0.166	0.093	0.813
Euler	0.650	0.356	0.125	0.192	0.029	0.937

**Table 6** Summary results by mathematical programming formulation

Form	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
Baseline	0.643	0.441	0.081	0.146	0.032	0.872
ExtraTol	0.716	0.410	0.141	0.203	0.028	0.911
SoftCons	0.637	0.356	0.092	0.141	0.026	0.904

#### 4.5 Additional disaggregations of the results

Appendix B studies different disaggregations of the above results, with the goal of getting further insights, not only of the individual elements involved in the different configurations, but also of potential interactions. They show that `daisy_mamil3F`, `harmonic` and `Lotka_VolterraF` are the easiest problems to solve and, moreover, that global optimization solvers seem to be better at finding the reference solution. Interestingly, Table 9 shows that the choice of the mathematical programming formulation can have a significant impact. `Baseline` performs extremely well in `hivF` and `daisy_mamil3F` problems, finding the reference solutions in more than 90% of the configurations, whereas the next best performing formulation for `hivF` is under 65%. On the other hand, `ExtraTol` is clearly the best formulation for `harmonic`, whereas `SoftCons` comes out narrowly on top for `Lotka_VolterraF`. Regarding the disaggregation by discretization scheme and mathematical programming formulation, we see in Table 12 that the configurations with `Trapezoid` tend to come out on top, while `Euler` and `Runge-Kutta` are at the bottom. Moreover, for all discretization schemes, the configurations with `Baseline` seem to dominate those with `ExtraTol`, which themselves seem to dominate those configurations with `SoftCons`.

#### 4.6 Running times by problem

Section 3 in the Online Appendix also provides detailed information about running times. In particular, it presents, for each problem, the top 10 configurations in terms of running times among those that have found the reference solution (top 10 among all solvers and also top 10 among global solvers).

The results show that `KnitroL` is the fastest solver in most of the problems. Some exceptions are `FHN`, where `IpopL`, `CONOPTL` and `BONMINL` are also quite fast, and `Lotka VolterraP`, `daisy mamil3P`, `CrausteP` and `CrausteF`, where again `IpopL` and `BONMINL` are relatively close to `KnitroL`. Regarding running times, local solvers are very fast, and `alpha pinene` is the only problem not solved by any configuration in less than a second (the top 10 fastest configurations go from 1.8 s to 8.9 s). Moreover, more than half of the problems are solved in less than 0.1 s by the fastest configuration.

If we restrict attention to global optimization solvers, the results are also quite promising. Setting again aside `alpha pinene`, for which no global solver was able to certify global optimality for the discretization at hand after 10 min, all other problems were solved to global optimality, by at least one configuration, in less than 10 s. `BARONG` is the fastest global optimization solver in most problems and, for some of them such as `BBG` and `hivP`, it is even competitive with the fastest local solver, `KnitroL`. On the other hand, `OcteractG` is the fastest in `Lotka VolterraF` and `CouenneG` is competitive with `BARONG` in `hivP`, `hivF` and `CrausteP`.

It may seem surprising that `alpha pinene`, despite coming from a relatively simple ODE system, has turned out to be relatively challenging, particularly for the certification of global optimality. Yet, recall that, differently from the other problems in the study, `alpha pinene` was fed with real (and noisy) measurements instead of synthetic ones. We have observed in some additional experiments with this problem that, with noiseless synthetic data, global certificates are consistently obtained within the time limits. The use of real measurements in this particular problem may have led to some practical lack of identifiability driven by the flatness of the objective function. These issues are discussed in the following section and, for the particular case of `alpha pinene`, they can be assessed in the figures in Appendix C.1.

## 5 Numerical results: local optimality, flatness, and identifiability

The approach followed in this paper, in which each problem is tackled with a wide range of configurations, including different families of local and global optimization algorithms, can be a useful tool to pinpoint structural characteristics of the underlying parameter estimation problems. In this section we illustrate the potential of our approach to accomplish this. We do so through a series of examples in which one can recognize patterns for multiplicity of local optima, flatness of the objective function, and lack of identifiability (i.e., multiplicity of global optima). Appendix C contains detailed results for all problems, similar to the selected ones that we discuss below.

### 5.1 Multiplicity of local optima

Overall, the existence of multiple local optima with significantly different objective functions (sum of squared errors in the resulting fit) does not seem to be an issue in the problems under study. In particular, the solutions provided by local optimizers are

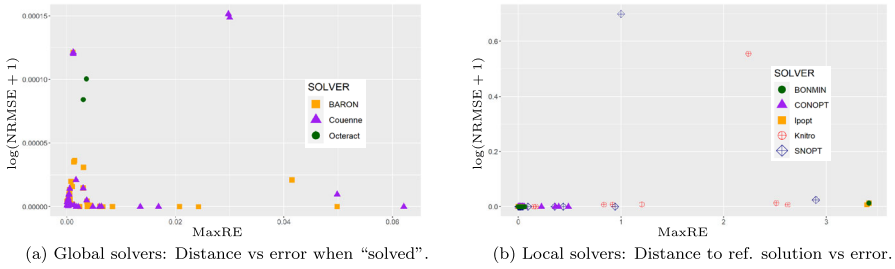


Fig. 1 Studying local optimality in `daisy_mam13P`

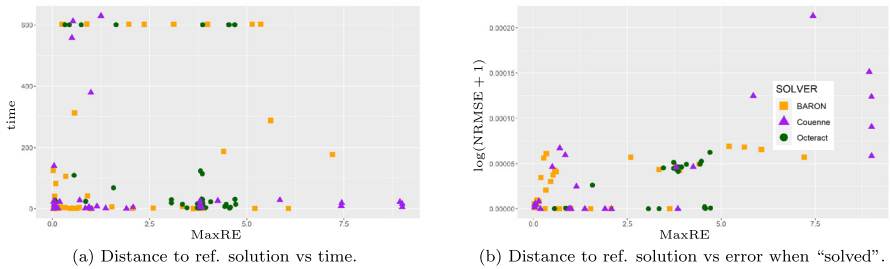


Fig. 2 Studying flatness in problem `hivP`

predominantly as close to the reference solution as the ones provided (and certified) by the global optimization solvers.

Figure 1 focuses on problem `daisy_mam13P`. The two plots represent the distance to the reference solution on the  $x$ -axis and the error on the  $y$ -axis, for the different configurations run for this problem. On the left, we represent the solutions provided by global solvers in configurations where they were able to certify global optimality on the discretized problem at hand. On the right we represent all the solutions provided by local solvers (just ensuring some local optimality condition).

We can see that, although the solutions provided by the global solvers tend to be closer to the reference solution, the ones provided by the local solvers are not far away and the associated errors are typically very small as well. In this case, the fact that the solutions of the local solvers are slightly more spread than the ones provided by the global solvers may be an indication of having a relatively flat objective function near the global optimum (which we discuss below), but there is no problem of getting stuck at local optima of very poor quality in terms of the error in the resulting fit.

### 5.2 Flatness of the objective function

If the objective function is very flat near the reference solution, convergence becomes harder and one might need to specify more demanding stopping criteria in order to get better solutions. One example of this can be seen in Fig. 2 for problem `hivP`.

The two plots represent, for global optimization solvers, the distance to the reference solution on the  $x$ -axis. On the left we have running time on the  $y$ -axis and, on the

right, the  $y$ -axis represents the error for the instances in which global optimality was “certified”. The plot of the left shows that most of the configurations successfully terminated within the time limit certifying global optimality of the final solution. Differently from the situation for `daisy_mamil3F` in Fig. 1, many of the obtained solutions are still relatively far from the reference solution. Yet, the plot on the right shows that all of these solutions still provide very small errors, and so imposing more strict termination criteria for the optimizers might mitigate this behavior for such a problem.

### 5.3 Lack of identifiability

Lack of identifiability can be an important challenge for parameter estimation problems. It arises when different parameter configurations deliver fits with the same quality. In other words, lack of identifiability corresponds to situations in which the underlying optimization problem has multiple global optima. It seems that lack of identifiability is not present in any of the problems we have studied since, as can be seen in all subfigures in Appendix C, it is never the case that global solvers provide very different solutions for a given problem beyond what we have classified above as flatness in the discussion of `hivP` in Fig. 2. This result is in agreement with the findings of (Go et al. 2023), where several of the problems considered here were shown to be globally identifiable.

## 6 Conclusions

The main goal of this paper was to study the potential of mathematical programming modeling and state-of-the-art solvers to handle well known parameter estimation problems. The results are very promising, significantly surpassing the capacity to solve these problems reported in past literature. Our study has also helped to assess different modeling aspects such as the chosen discretization scheme or different variants on the formulation of the different optimization problems.

A natural direction for future research is to explore the robustness and scalability of the analysis, systematically introducing noisy measurements and studying significantly larger problems. In this respect, it is worth emphasizing that in our numerical analysis we have not relied on any sort of parallelization, which could definitely be exploited by solvers with these kind of capabilities such as `OcateractG` and `KnitroL`. Further, it is also worth noting that all the analysis developed in this paper has been done using the default settings of each solver and, thus, fine tuning these settings for each pair solver-problem might enable the solution of significantly larger problems. A related approach would be to develop a methodology for identifying the most promising configurations for solving a given problem, such as the choice of discretization scheme, mathematical programming formulation, or solver.

The analysis we have developed here can also serve as a benchmark for evaluating the performance of new specialized algorithms, specially global optimization

ones, since they can now be compared to solvers such as Couenne<sup>G</sup>, BARON<sup>G</sup>, and Octeract<sup>G</sup>.

## Appendix A. Detailed description of the problems

This appendix contains the detailed description of the parameter estimation problems used in the numerical experiments.

### Problem: alpha pinene

ODE system:

$$\begin{aligned}\frac{dx_1}{dt} &= -(p_1 + p_2)x_1 \\ \frac{dx_2}{dt} &= p_1x_1 \\ \frac{dx_3}{dt} &= p_2x_1 - (p_3 + p_4)x_3 + p_5x_5 \\ \frac{dx_4}{dt} &= p_3x_3 \\ \frac{dx_5}{dt} &= -p_4x_3 + p_5x_5\end{aligned}$$

Initial conditions:  $x_1(0) = 100$ ,  $x_2(0) = x_3(0) = x_4(0) = x_5(0) = 0$ .

Observed states:  $y_i(t) = x_i(t)$ ,  $i = 1, \dots, 5$ .

Measurements for each observed variable: The time interval considered is  $[0, 36900]$ , with 8 measurements at time points 1230, 3060, 4920, 7800, 10680, 15030, 22620, and 36420.

Parameter bounds:  $0 \leq p_i \leq 1$ ,  $i = 1, \dots, 5$ .

### Problem: BBG

ODE system:

$$\begin{aligned}\frac{dC_b}{dt} &= \mu_{max} \frac{C_s C_b}{K_s + C_s} - k_d C_b \\ \frac{dC_s}{dt} &= -\frac{\mu_{max}}{yield} \frac{C_s C_b}{K_s + C_s}\end{aligned}$$

Initial conditions:  $C_b(0) = 2$  and  $C_s(0) = 30$ .

Observed states:  $y_1(t) = C_b(t)$  and  $y_2(t) = C_s(t)$ .

Measurements for each observed variable: 7 measurements uniformly distributed in the interval  $[0, 12]$ .

Parameter bounds:  $0.0001 \leq p_i \leq 100$ ,  $i = 1, \dots, 4$ , where  $(p_1, p_2, p_3, p_4) = (\mu_{max}, K_s, k_d, yield)$ .

### Problem: FHN

ODE system:

$$\begin{aligned}\frac{dV}{dt} &= g \left( V - \frac{V^3}{3} + R \right) \\ \frac{dR}{dt} &= \frac{1}{g} (V - a + bR)\end{aligned}$$

Initial conditions:  $V(0) = -1$  and  $R(0) = 1$ .

Observed states:  $y_1(t) = V(t)$ .

Measurements for each observed variable: 6 measurements uniformly distributed in the interval  $[0, 20]$ .

Parameter bounds:  $10^{-5} \leq p_i \leq 10^5$ ,  $i = 1, \dots, 3$ , where  $(p_1, p_2, p_3) = (g, a, b)$ .

### Problem: harmonic

ODE system:

$$\begin{aligned}\frac{dx_1}{dt} &= -p_1 x_2 \\ \frac{dx_2}{dt} &= \frac{1}{p_2} x_1\end{aligned}$$

Initial conditions:  $0 \leq x_i(0) \leq 1.5$ ,  $i = 1, 2$ .

Observed states:  $y_i(t) = x_i(t)$ ,  $i = 1, 2$ .

Measurements for each observed variable: 10 measurements uniformly distributed in the interval  $[0, 2.3]$ .

Parameter bounds:  $0.0001 \leq p_i \leq 10$ ,  $i = 1, 2$ .

### Problem: Lotka Volterra

ODE system:

$$\begin{aligned}\frac{dr}{dt} &= k_1 r - k_2 r w \\ \frac{dw}{dt} &= k_2 r w - k_3 w\end{aligned}$$

Initial conditions:  $90 \leq r(0) \leq 110$  and  $90 \leq w(0) \leq 110$ .

Observed states:

$$\begin{array}{l|l} \text{Lotka Volterra}^{\text{F}} & \text{Lotka Volterra}^{\text{P}} \\ y_1(t) = r(t) & y_1(t) = r(t) \\ y_2(t) = w(t) & \end{array}$$

Measurements for each observed variable: 20 measurements uniformly distributed in the interval  $[0, 1]$ .

Parameter bounds:  $0.0001 \leq p_i \leq 1, i = 1, \dots, 3$ , where  $(p_1, p_2, p_3) = (k_1, k_2, k_3)$ .

### Problem: daisy mamil

ODE system:

$$\begin{aligned} \frac{dx_1}{dt} &= -(a_{21} + a_{31} + a_{01})x_1 + a_{12}x_2 + a_{13}x_3 \\ \frac{dx_2}{dt} &= a_{21}x_1 - a_{12}x_2 \\ \frac{dx_3}{dt} &= a_{31}x_1 - a_{13}x_3 \end{aligned}$$

Initial conditions:

$$\begin{aligned} -1 &\leq x_1(0) \leq 2 \\ -1 &\leq x_2(0) \leq 2 \\ -1 &\leq x_3(0) \leq 2 \end{aligned}$$

Observed states:

$$\begin{array}{l|l} \text{daisy mamil3}^{\text{F}} & \text{daisy mamil3}^{\text{P}} \\ y_1(t) = x_1(t) & y_1(t) = x_1(t) \\ y_2(t) = x_2(t) & y_2(t) = x_2(t) \\ y_3(t) = x_3(t) & \end{array}$$

Measurements for each observed variable: 20 measurements uniformly distributed in the interval  $[0, 1]$ .

Parameter bounds:  $-1 \leq p_i \leq 2, i = 1, \dots, 5$ , where  $(p_1, p_2, p_3, p_4, p_5) = (a_{21}, a_{31}, a_{01}, a_{12}, a_{13})$ .

**Problem: hiv**

ODE system:

$$\begin{aligned}\frac{dx}{dt} &= \lambda - dx - \beta xv \\ \frac{dy}{dt} &= \beta xv - ay \\ \frac{dv}{dt} &= ky - uv \\ \frac{dw}{dt} &= cxyw - cqyw - bw \\ \frac{dz}{dt} &= cqyw - hz\end{aligned}$$

Initial conditions:

$$\begin{aligned}0.001 &\leq x(0) \leq 2 \\ 0.001 &\leq y(0) \leq 2 \\ 0.001 &\leq v(0) \leq 2 \\ 0.001 &\leq w(0) \leq 2 \\ 0.001 &\leq z(0) \leq 2\end{aligned}$$

Observed states:

$$\begin{array}{l|l}\text{hiv}^F & \text{hiv}^P \\ y_1(t) = x(t) & y_1(t) = x(t) \\ y_2(t) = y(t) & y_2(t) = y(t) + v(t) \\ y_3(t) = v(t) & y_3(t) = w(t) \\ y_4(t) = w(t) & y_4(t) = z(t) \\ y_5(t) = z(t) & \end{array}$$

Measurements for each observed variable: 20 measurements uniformly distributed in the interval  $[0, 10]$ .

Parameter bounds:  $0.0001 \leq p_i \leq 1$ ,  $i = 1, \dots, 10$ , where  $(p_1, p_2, \dots, p_{10}) = (\lambda, d, \dots, h)$ .

**Problem: Crauste**

ODE system:

$$\frac{dN}{dt} = \mu_N N - \delta_{NE} N P$$

$$\begin{aligned} \frac{dE}{dt} &= \delta_{NE}NP - \mu_{EE}E^2 - \delta_{EL}E + \rho_{EE}EP \\ \frac{dS}{dt} &= \delta_{EL}S - S\delta_{LM} - \mu_{LL}S^2 - \mu_{LE}ES \\ \frac{dM}{dt} &= \delta_{LM}S - \mu_{MM}M \\ \frac{dP}{dt} &= \rho_{PP}P^2 - \mu_{PP}P - \mu_{PE}EP - \mu_{PL}SP \end{aligned}$$

Initial conditions:

$$\begin{aligned} -1.1 &\leq N(0) \leq 1.1 \\ -1.1 &\leq E(0) \leq 1.1 \\ -1.1 &\leq S(0) \leq 1.1 \\ -1.1 &\leq M(0) \leq 1.1 \\ -1.1 &\leq P(0) \leq 1.1 \end{aligned}$$

Observed states:

Crauste <sup>F</sup>	Crauste <sup>P</sup>
$y_1(t) = N(t)$	$y_1(t) = N(t)$
$y_2(t) = E(t)$	$y_2(t) = E(t)$
$y_3(t) = S(t)$	$y_3(t) = S(t) + M(t)$
$y_4(t) = M(t)$	$y_4(t) = P(t)$
$y_5(t) = P(t)$	

Measurements for each observed variable: 20 measurements uniformly distributed in the interval  $[0, 1]$ .

Parameter bounds:  $-2 \leq p_i \leq 2, i = 1, \dots, 13$ , where  $(p_1, p_2, p_3, \dots, p_{13}) = (\mu_N, \delta_{NE}, \mu_{EE}, \dots, \mu_{PL})$ .

## Appendix B. Numerical results: digging deeper

The goal of this section is to get additional insights from the computational analysis by studying different disaggregations of the numerical results. The objective here is to get a better understanding, not only of the individual elements involved in the different configurations, but also of any potential interactions between them. For the sake of exposition, here we only present the top performing configurations, with respect to FOUND<sup>R</sup>, for each considered disaggregation level.<sup>10</sup>

The results highlight once again that there is no need to fine tune a specific configuration to be able to successfully solve a given problem, since in most of them the reference solution is solved with a wide variety of solvers and configurations.

<sup>10</sup> The full tables are available in the Online Appendix.

**Table 7** Summary results by problem and solver

Problem — solver	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
daisy mamil3 <sup>F</sup> — BARON <sup>G</sup>	0.880	1.000	0	0	0.076	1.000
daisy mamil3 <sup>F</sup> — Couenne <sup>G</sup>	1.000	1.000	0	0	0.220	1.000
harmonic — Oocteract <sup>G</sup>	0.720	0.980	0.020	0	1.286	1.000
harmonic — BARON <sup>G</sup>	0.760	0.960	0	0	0.143	1.000
daisy mamil3 <sup>F</sup> — BONMIN <sup>L</sup>	1.000	0.960	0	0	0.046	1.000
daisy mamil3 <sup>F</sup> — CONOPT <sup>L</sup>	0	0.920	0.080	0	0.072	1.000
daisy mamil3 <sup>F</sup> — Ipopt <sup>L</sup>	1.000	0.920	0	0	0.056	1.000
Lotka Volterra <sup>F</sup> — BONMIN <sup>L</sup>	1.000	0.900	0	0	0.038	1.000
Lotka Volterra <sup>F</sup> — Ipopt <sup>L</sup>	0.980	0.880	0	0	0.038	0.980
Lotka Volterra <sup>F</sup> — BARON <sup>G</sup>	0.580	0.860	0	0	0.097	1.000

**Table 8** Summary results by problem and discretization scheme

Problem — scheme	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
Lotka Volterra <sup>F</sup> — Simpson	0.775	0.988	0.012	0	0.028	1.000
Lotka Volterra <sup>F</sup> — Trapezoid	0.775	0.975	0.025	0	0.026	1.000
daisy mamil3 <sup>F</sup> — Euler	0.700	0.925	0.025	0	0.029	1.000
daisy mamil3 <sup>F</sup> — Trapezoid	0.750	0.925	0.025	0	0.031	1.000
Lotka Volterra <sup>F</sup> — Adams-Moulton	0.863	0.900	0.012	0	0.029	0.912
daisy mamil3 <sup>F</sup> — Simpson	0.725	0.887	0.012	0	0.038	0.938
daisy mamil3 <sup>F</sup> — Runge-Kutta	0.750	0.875	0.025	0.012	0.093	0.938
Lotka Volterra <sup>F</sup> — Runge-Kutta	0.725	0.825	0.050	0	0.102	0.900
daisy mamil3 <sup>F</sup> — Adams-Moulton	0.750	0.812	0.025	0	0.055	0.950
harmonic — Trapezoid	0.787	0.800	0	0	0.039	0.950

Table 7 confirms that daisy mamil3<sup>F</sup>, harmonic and Lotka Volterra<sup>F</sup> are the easiest problems to solve and, moreover, that global optimization solvers seem to be better at finding the reference solution.

Table 8 shows that, although Euler seems to be the worst performing discretization scheme overall, it can be particularly effective for specific problems. In particular, it is the best performing scheme in daisy mamil3<sup>F</sup>, tied with Trapezoid.

Interestingly, Table 9 shows that the choice of the mathematical programming formulation can have a significant impact. Baseline performs extremely well in hiv<sup>F</sup> and daisy mamil3<sup>F</sup> problems, finding the reference solutions in more than 90% of the configurations, whereas the next best performing formulation for hiv<sup>F</sup> is under 65%. On the other hand, ExtraTol is clearly the best formulation for harmonic, whereas SoftCons comes out narrowly on top for Lotka Volterra<sup>F</sup>.<sup>11</sup>

<sup>11</sup> The Online Appendix contains the full tables.

**Table 9** Summary results by problem and mathematical programming formulation

Problem — form	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
daisy mamil3 <sup>F</sup> — Baseline	0.725	0.938	0	0	0.032	0.975
hiv <sup>F</sup> — Baseline	0.838	0.925	0	0	0.056	0.938
daisy mamil3 <sup>F</sup> — ExtraTol	0.738	0.875	0.056	0.006	0.029	0.956
daisy mamil3 <sup>F</sup> — SoftCons	0.738	0.869	0	0	0.040	0.969
harmonic — ExtraTol	0.812	0.863	0	0	0.034	0.981
Lotka Volterra <sup>F</sup> — SoftCons	0.738	0.844	0.037	0	0.026	0.975
Lotka Volterra <sup>F</sup> — ExtraTol	0.769	0.838	0.019	0	0.030	0.963
Lotka Volterra <sup>F</sup> — Baseline	0.700	0.812	0	0	0.044	0.925
harmonic — SoftCons	0.619	0.650	0	0	0.120	0.894
hiv <sup>F</sup> — ExtraTol	0.856	0.644	0.150	0.206	0.033	1.000

**Table 10** Summary results by solver and discretization scheme

Solver — scheme	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
Couenne <sup>G</sup> — Trapezoid	0.795	0.591	0.144	0.098	0.138	0.917
BARON <sup>G</sup> — Trapezoid	0.758	0.576	0.114	0.182	0.078	1.000
BARON <sup>G</sup> — Adams-Moulton	0.758	0.568	0.129	0.144	0.102	0.985
Knitro <sup>L</sup> — Trapezoid	0.977	0.530	0.098	0.174	0.026	1.000
Ocateract <sup>G</sup> — Trapezoid	0.568	0.523	0.106	0.167	1.286	0.939
Couenne <sup>G</sup> — Simpson	0.780	0.523	0.182	0.083	0.148	0.894
BARON <sup>G</sup> — Simpson	0.689	0.500	0.159	0.167	0.095	1.000
BARON <sup>G</sup> — Runge-Kutta	0.667	0.485	0.121	0.136	0.306	0.977
Couenne <sup>G</sup> — Adams-Moulton	0.720	0.477	0.121	0.106	0.236	0.924
BARON <sup>G</sup> — Euler	0.462	0.462	0.152	0.212	0.076	1.000

Table 10 shows that Trapezoid and is the best performing configuration for global optimization solvers solvers.

Table 11 shows that Baseline is often the best performing formulation for the different optimization solvers, although ExtraTol comes out on top for BARON<sup>G</sup> and Ipopt<sup>L</sup>, for instance.

Finally, regarding the disaggregation by discretization scheme and mathematical programming formulation, we see in Table 12 that the configurations with Trapezoid tend to come out on top and Euler and Runge-Kutta are at the bottom. Moreover, for all discretization schemes, the configurations with Baseline seem to dominate those with ExtraTol, which themselves seem to dominate those configurations with SoftCons.

**Table 11** Summary results by solver and mathematical programming formulation

Solver — form	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
Knitro <sup>L</sup> — Baseline	0.888	0.576	0.096	0.128	0.032	0.984
BARON <sup>G</sup> — ExtraTol	0.786	0.572	0.182	0.133	0.076	0.996
Couenne <sup>G</sup> — Baseline	0.664	0.536	0.096	0.040	0.212	0.784
BONMIN <sup>L</sup> — Baseline	0.856	0.504	0.120	0.024	0.046	0.856
Couenne <sup>G</sup> — ExtraTol	0.705	0.491	0.158	0.109	0.138	0.849
BARON <sup>G</sup> — Baseline	0.624	0.488	0.096	0.264	0.134	1.000
BARON <sup>G</sup> — SoftCons	0.552	0.472	0.100	0.160	0.097	0.984
Ocateract <sup>G</sup> — Baseline	0.616	0.456	0.056	0.176	1.563	0.856
Knitro <sup>L</sup> — ExtraTol	0.975	0.449	0.168	0.186	0.028	0.993
Ipopt <sup>L</sup> — ExtraTol	0.902	0.449	0.154	0.175	0.056	0.937

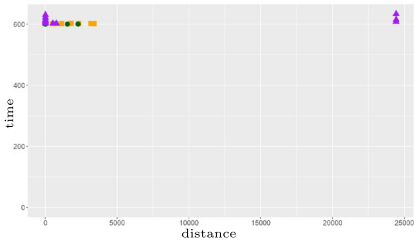
**Table 12** Summary results by discretization scheme and mathematical programming formulation

Scheme — form	SOLVED <sup>S</sup>	FOUND <sup>R</sup>	NEAR <sup>R</sup>	ALTERN	TIME <sup>BFR</sup>	SUCCESS
Trapezoid — Baseline	0.715	0.530	0.065	0.145	0.032	0.950
Trapezoid — ExtraTol	0.750	0.474	0.125	0.206	0.028	0.950
Adams-Moulton — Baseline	0.700	0.445	0.075	0.145	0.093	0.870
Simpson — Baseline	0.635	0.430	0.105	0.130	0.046	0.870
Adams-Moulton — ExtraTol	0.724	0.425	0.149	0.189	0.041	0.897
Simpson — ExtraTol	0.741	0.419	0.140	0.193	0.037	0.925
Trapezoid — SoftCons	0.695	0.415	0.100	0.160	0.026	0.953
Runge-Kutta — Baseline	0.545	0.405	0.055	0.140	0.106	0.760
Euler — Baseline	0.620	0.395	0.105	0.170	0.032	0.910
Runge-Kutta — ExtraTol	0.664	0.371	0.138	0.189	0.093	0.840

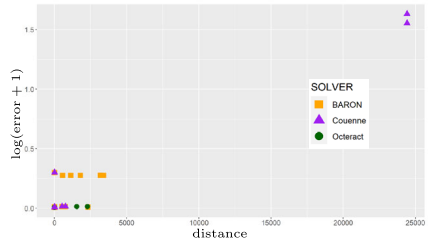
## Appendix C. Graphical representations of the results

### Appendix C.1. `alpha_pinene`

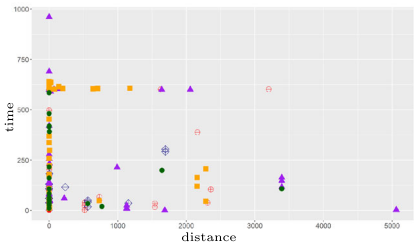
See Fig. 3



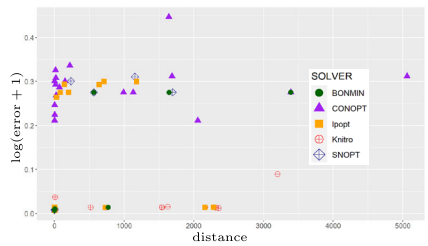
(a) Global solvers: Distance to ref. solution vs time.



(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.

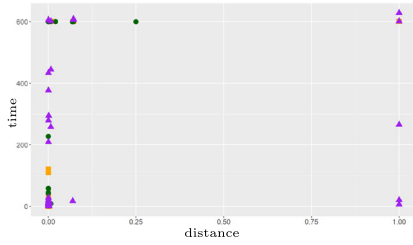


(d) Local solvers: Distance to ref. solution vs error.

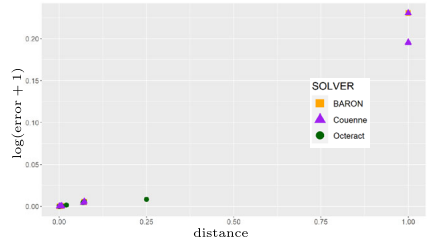
**Fig. 3** Results for problem `alpha_pinene`

Appendix C.2. harmonic

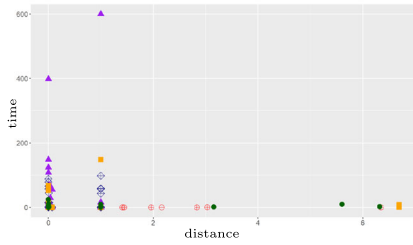
See Fig. 4



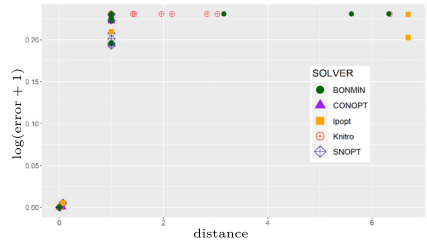
(a) Global solvers: Distance to ref. solution vs time.



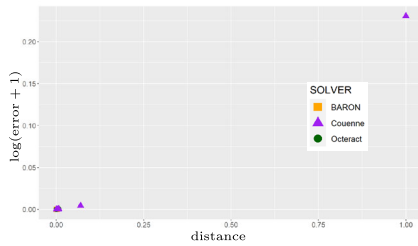
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

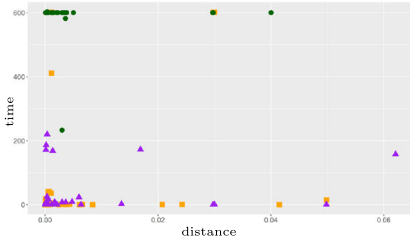


(e) Global solvers: Distance to ref. solution vs error when "solved".

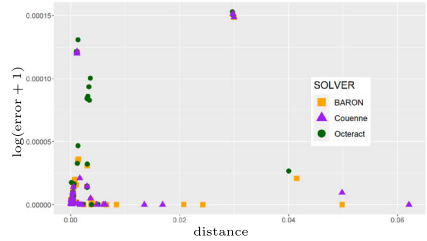
Fig. 4 Results for problem harmonic

Appendix C.3. *daisy* *mami13*<sup>F</sup>

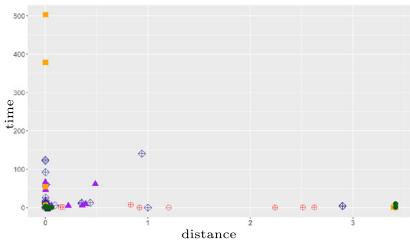
See Fig. 5



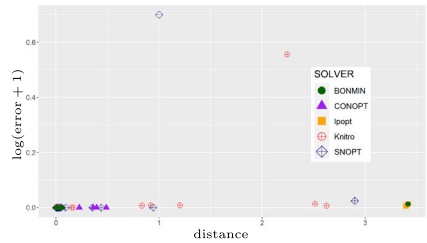
(a) Global solvers: Distance to ref. solution vs time.



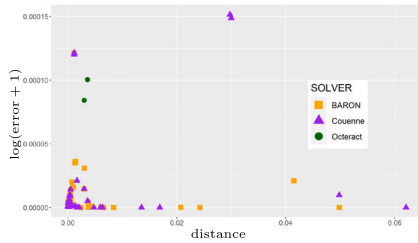
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

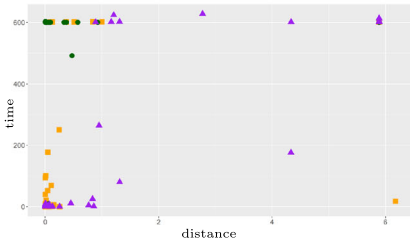


(e) Global solvers: Distance to ref. solution vs error when "solved".

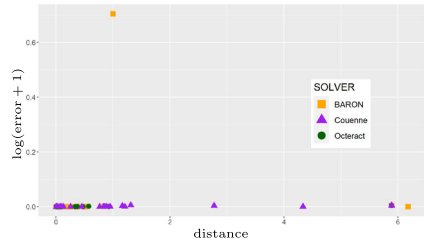
Fig. 5 Results for problem *daisy* *mami13*<sup>F</sup>

Appendix C.4. *daisy mam13*<sup>P</sup>

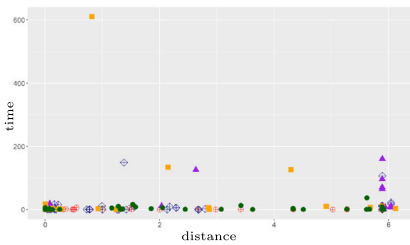
See Fig. 6



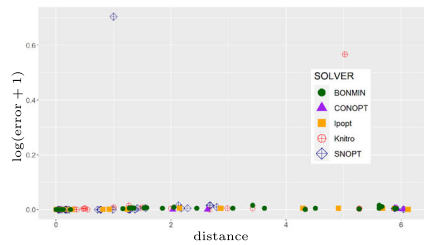
(a) Global solvers: Distance to ref. solution vs time.



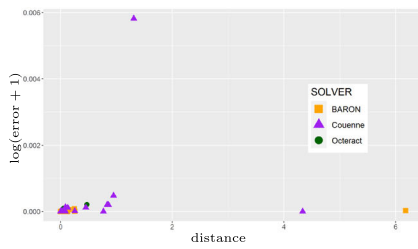
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

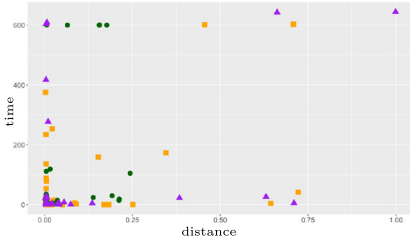


(e) Global solvers: Distance to ref. solution vs error when "solved".

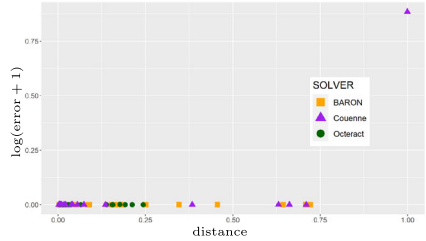
Fig. 6 Results for problem *daisy mam13*<sup>P</sup>

**Appendix C.5. hiv<sup>F</sup>**

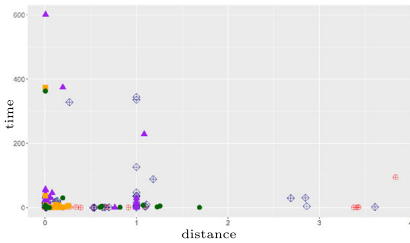
See Fig. 7



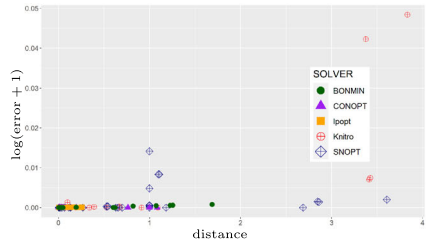
(a) Global solvers: Distance to ref. solution vs time.



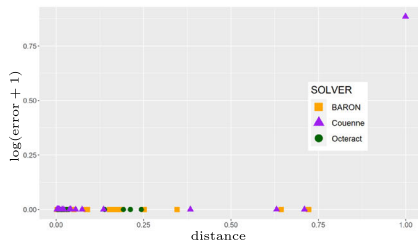
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

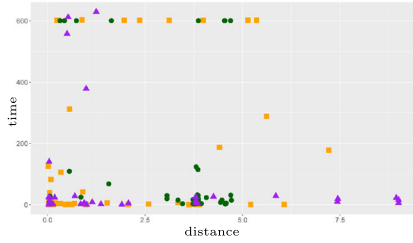


(e) Global solvers: Distance to ref. solution vs error when "solved".

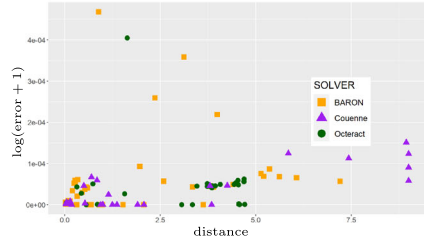
**Fig. 7** Results for problem hiv<sup>F</sup>

Appendix C.6.  $hiv^P$

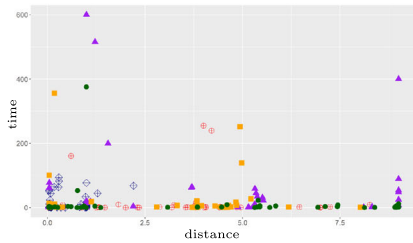
See Fig. 8



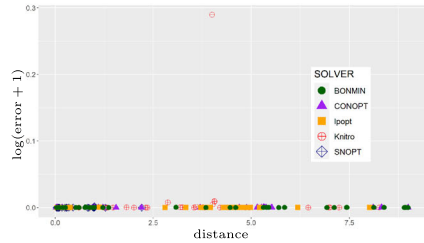
(a) Global solvers: Distance to ref. solution vs time.



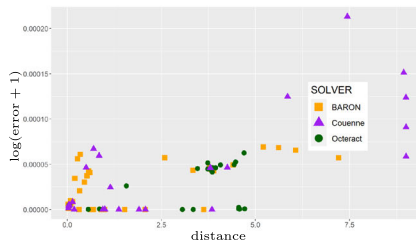
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

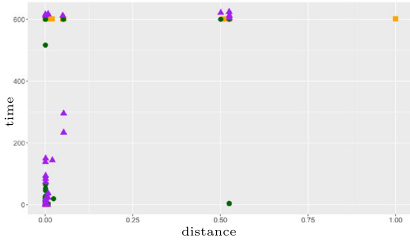


(e) Global solvers: Distance to ref. solution vs error when "solved".

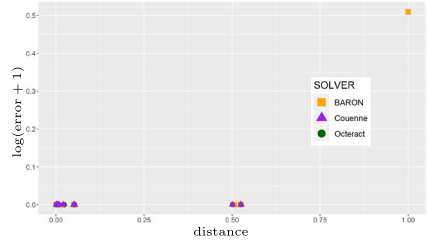
Fig. 8 Results for problem  $hiv^P$

### Appendix C.7. Lotka Volterra<sup>F</sup>

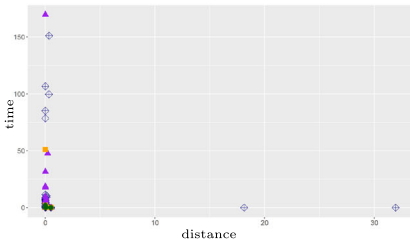
See Fig. 9



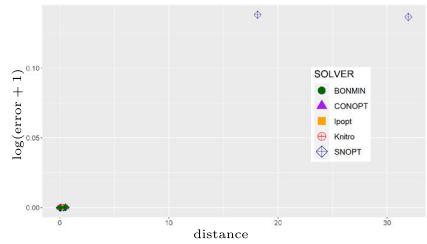
(a) Global solvers: Distance to ref. solution vs time.



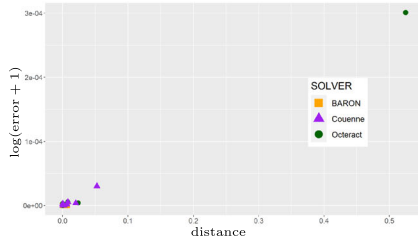
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

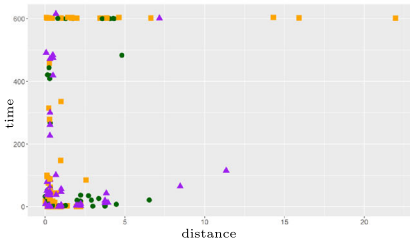


(e) Global solvers: Distance to ref. solution vs error when "solved".

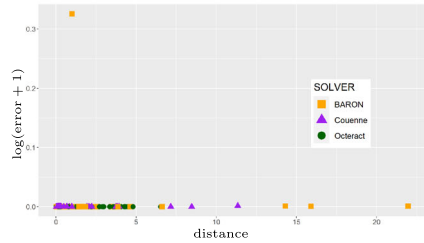
Fig. 9 Results for problem Lotka Volterra<sup>F</sup>

### Appendix C.8. Lotka Volterra<sup>P</sup>

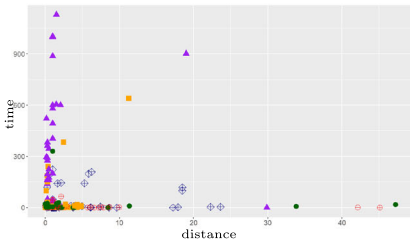
See Fig. 10



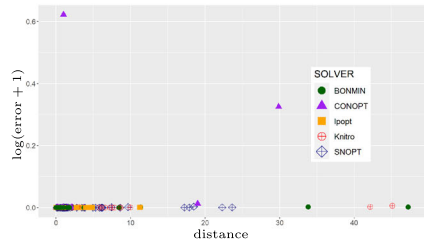
(a) Global solvers: Distance to ref. solution vs time.



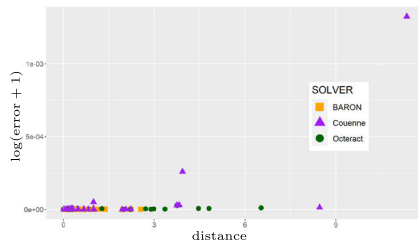
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

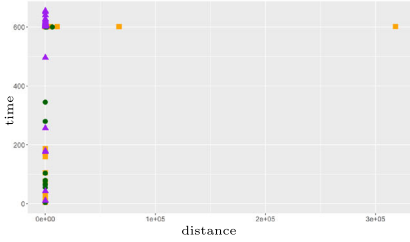


(e) Global solvers: Distance to ref. solution vs error when "solved".

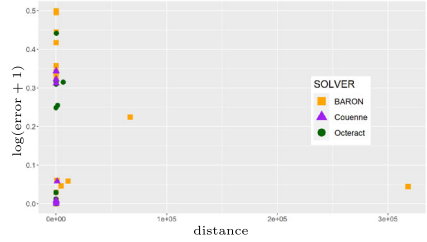
Fig. 10 Results for problem Lotka Volterra<sup>P</sup>

**Appendix C.9. FHN**

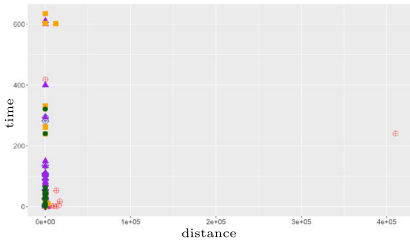
See Fig. 11



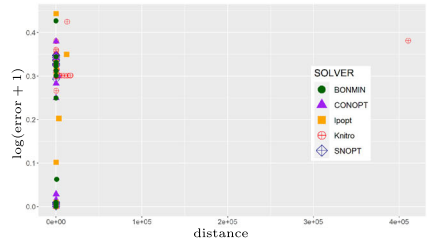
(a) Global solvers: Distance to ref. solution vs time.



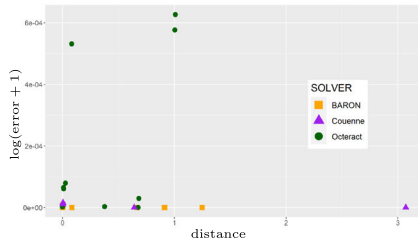
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

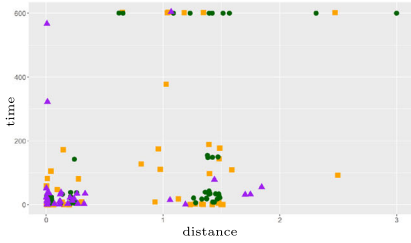


(e) Global solvers: Distance to ref. solution vs error when "solved".

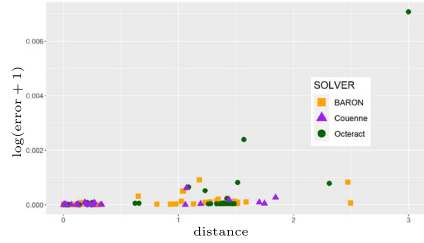
**Fig. 11** Results for problem FHN

### Appendix C.10. Crauste<sup>F</sup>

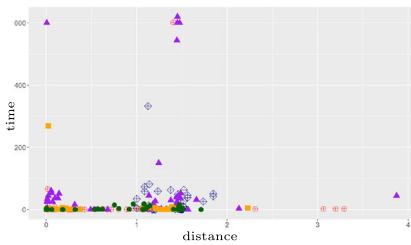
See Fig. 12



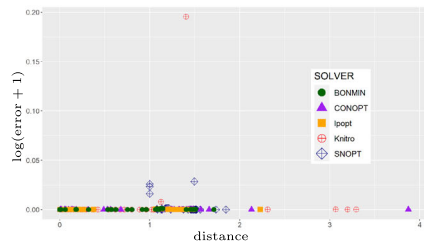
(a) Global solvers: Distance to ref. solution vs time.



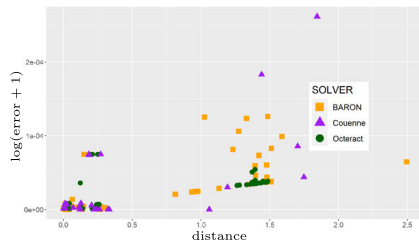
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

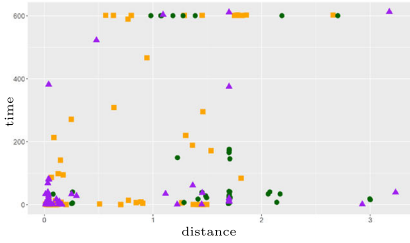


(e) Global solvers: Distance to ref. solution vs error when "solved".

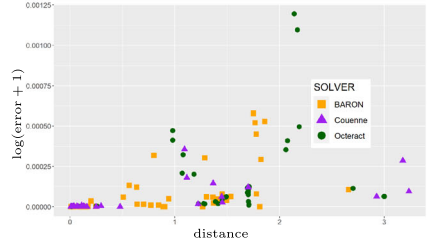
**Fig. 12** Results for problem Crauste<sup>F</sup>

### Appendix C.11. Crauste<sup>P</sup>

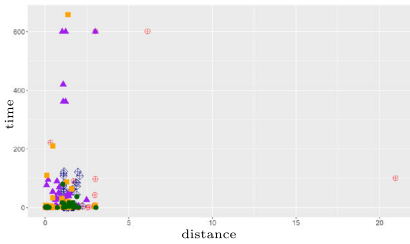
See Fig. 13



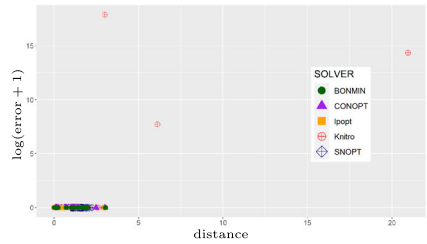
(a) Global solvers: Distance to ref. solution vs time.



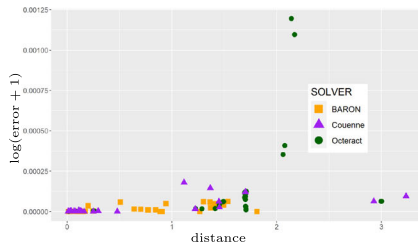
(b) Global solvers: Distance to ref. solution vs error.



(c) Local solvers: Distance to ref. solution vs time.



(d) Local solvers: Distance to ref. solution vs error.

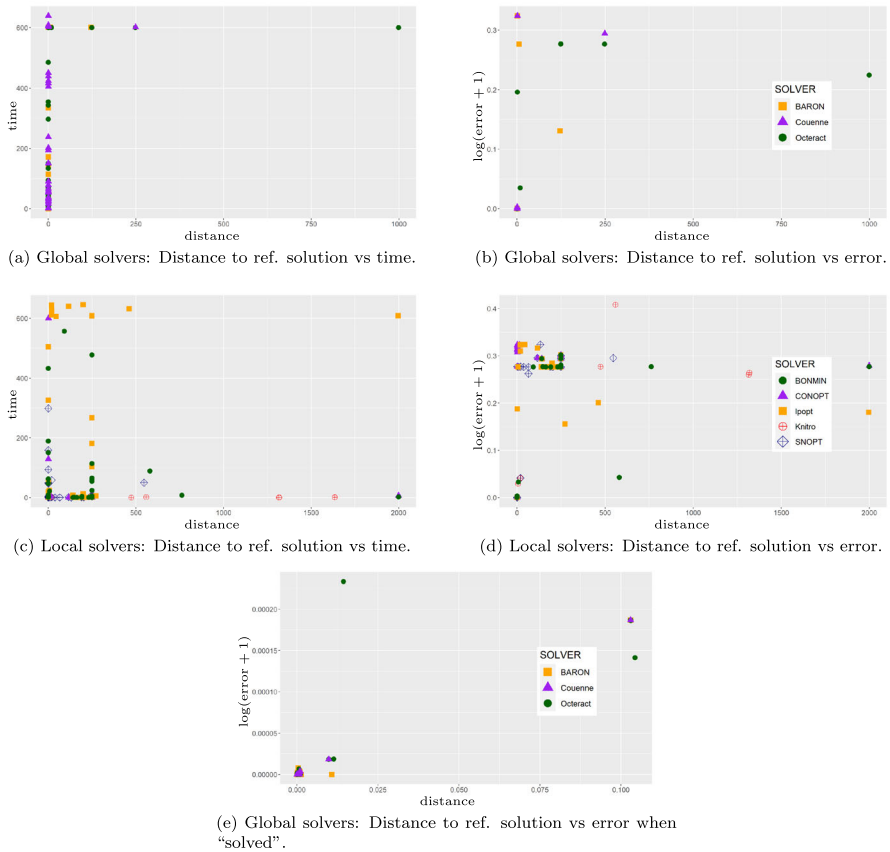


(e) Global solvers: Distance to ref. solution vs error when "solved".

Fig. 13 Results for problem Crauste<sup>P</sup>

## Appendix C.12. BGG

See Fig. 14



**Fig. 14** Results for problem BGG

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11081-025-09978-9>.

**Acknowledgements** The authors thank Eliseo Pita Vilarriño for his contribution to the automation of the filtering and the processing of the computational results.

**Author Contributions** M.F-D implemented the different solution methodologies as well as their adjustment to the different problems under study. M.F-D also carried out the computational study and collected the numerical results. All five authors participated in the selection of the target problems, the design of the mathematical programming formulations, the design of the computational study, and the analysis, interpretation and discussion of the numerical results.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work is part of the R&D projects PID2021-124030NB-C31 and PID2021-124030NB-C32 funded by MICIU/AEI/10.13039/501100011033/ and by ERDF/EU. This research was also funded by Grupos de Referencia Competitiva ED431C-2021/24 from the Consellería de Cultura, Educación e Universidades, Xunta de Galicia. JRB acknowledges support from grant PID2020-117271RB-C22 (BIODYNAMICS) funded by MCIN/AEI/10.13039/501100011033, from grant PID2023-146275NB-C22 (DYNAMO-bio) funded by MICIU/AEI/ 10.13039/501100011033 and ERDF/EU, and from grant CSIC PIE 202470E108 (LARGO). The authors acknowledge CESGA (Centro de Supercomputación de Galicia) for providing access to its FinisTerra III supercomputer.

**Availability of data and materials** Upon acceptance, all AMPL models used in the computational study will be made publicly available according to the journal standards.

## Declarations

**Conflict of interest** The authors declare that they have no Conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG (2009) Systems biology: parameter estimation for biochemical models. *FEBS J* 276:886–902
- Balsa-Canto E, Peifer M, Banga JR, Timmer J, Fleck C (2008) Hybrid optimization method with general switching strategy for parameter estimation. *BMC Syst Biol* 2:1–9
- Banga JR, Balsa-Canto E (2008) Parameter estimation and optimal experimental design. *Essays Biochem* 45:195–210
- Bellu G, Saccomani MP, Audoly S, D'Angiò L (2007) Daisy: A new software tool to test global identifiability of biological and physiological systems. *Comput Methods Programs Biomed* 88:52–61
- Belotti P, Lee J, Liberti L, Margot F, Wächter A (2009) Branching and bounds tightening techniques for non-convex MINLP. *Optimiz Methods Software* 24:597–634
- Betts JT (2020) Practical methods for optimal control using nonlinear programming, 3rd edn. Society for Industrial and Applied Mathematics, Philadelphia
- Biegler LT (2007) An overview of simultaneous strategies for dynamic optimization. *Chem Eng Process* 46:1043–1053
- Biegler LT (2010) Nonlinear programming: concepts, algorithms, and applications to chemical processes. SIAM
- Bonami P, Kilinç M, Linderoth J (2012) Algorithms and software for convex mixed integer nonlinear programs. In: Lee J, Leyffer S (eds) *Mixed Integer Nonlinear Programming*. Springer, New York, pp 1–39
- Borchers HW (2023) *pracma: Practical Numerical Math Functions*. <https://CRAN.R-project.org/package=pracma>. r package version 2.4.4
- Bryson AE, Ho YC (1969) *Applied Optimal Control*. Blaisdell
- Burden RL, Faires JD (2001) *Numerical analysis*, 7th edn. Prindle Weber and Schmidt, Boston
- Byrd RH, Nocedal J, Waltz RA (2006) *Knitro: an integrated package for nonlinear optimization*
- Chachuat B, Singer AB, Barton PI (2006) Global methods for dynamic optimization and mixed-integer dynamic optimization. *Indust Eng Chem Res* 45:8373–8392

- Chen WW, Niepel M, Sorger PK (2010) Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev* 24:1861–1875
- Chou IC, Voit EO (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math Biosci* 219:57–83
- Crauste F, Mañilla J, Boucinha L, Djebali S, Gandrillon O, Marvel J, Arpin C (2017) Identification of nascent memory CD8 T cells and modeling of their ontogeny. *Cell Syst* 4:306–317
- Czyzyk J, Mesnier MP, Moré JJ (1998) The NEOS Server. *IEEE J Comput Sci Eng* 5:68–75
- Drud A (1985) CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems. *Math Program* 31:153–191
- Edsberg L, Wedin PÅ (1995) Numerical tools for parameter estimation in ode-systems. *Optimiz Methods Software* 6:193–217
- Egea JA, Martí R, Banga JR (2010) An evolutionary method for complex-process optimization. *Comput Oper Res* 37:315–324
- Esposito WR, Floudas CA (2000) Global optimization for the parameter estimation of differential-algebraic systems. *Ind Eng Chem Res* 39:1291–1310
- FitzHugh R (1961) Impulses and physiological states in theoretical models of nerve membrane. *Biophys J* 1:445–466
- Fourer R, Gay DM, Kernighan BW (1990) AMPL: a mathematical programming language. *Manage Sci* 36:519–554
- Fröhlich F, Kaltenbacher B, Theis FJ, Hasenauer J (2017) Scalable parameter estimation for genome-scale biochemical reaction networks. *PLoS Comput Biol* 13:e1005331
- Fuguitt RE, Hawkins JE (1945) The liquid phase thermal isomerization of  $\alpha$ -pinene 1a, 1b. *J Am Chem Soc* 67:242–245
- Gill PE, Murray W, Saunders MA (2005) SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev* 47:99–131
- Go S, Hong H, Ilmer I, Ovchinnikov A, Soto P, Yap C (2023) Symbolic-numeric parameter estimation software package in Julia. arXiv 2303.02159v1
- Goh C, Teo KL (1988) Control parametrization: a unified approach to optimal control problems with general constraints. *Automatica* 24:3–18
- Jia G, Stephanopoulos G, Gunawan R (2012) Incremental parameter estimation of kinetic metabolic network models. *BMC Syst Biol* 6:1–12
- Lin Y, Stadtherr MA (2006) Deterministic global optimization for parameter estimation of dynamic systems. *Ind Eng Chem Res* 45:8438–8448
- Lin Y, Stadtherr MA (2007) Deterministic global optimization of nonlinear dynamic systems. *AICHE J* 53:866–875
- Ljung L, Chen T (2013) Convexity issues in system identification. In: 2013 10th IEEE International Conference on Control and Automation (ICCA), IEEE
- Mendes P, Kell D (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinform* (Oxford, England) 14:869–883
- Miró A, Pozo C, Guillén-Gosálbez G, Egea JA, Jiménez L (2012) Deterministic global optimization algorithm based on outer approximation for the parameter estimation of nonlinear dynamic biological systems. *BMC Bioinformatics* 13
- Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res* 13:2467–2474
- Nagumo J, Arimoto S, Yoshizawa S (1962) An active pulse transmission line simulating nerve axon. *Proc IRE* 50:2061–2070
- Octeract Optimisation Intelligence (2023) Octeract Engine
- Panning TD, Watson LT, Allen NA, Chen KC, Shaffer CA, Tyson JJ (2008) Deterministic parallel global parameter estimation for a model of the budding yeast cell cycle. *J Global Optim* 40:719–738
- Papamichail I, Adjiman C (2004) Global optimization of dynamic systems. *Comput Chem Eng* 28:403–415
- Papamichail I, Adjiman CS (2002) A rigorous global optimization algorithm for problems with ordinary differential equations. *J Global Optim* 24:1–33
- Penas DR, González P, Egea JA, Doallo R, Banga JR (2017) Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy. *BMC Bioinformatics*, 18
- Pérez-Galván C, Bogle IDL (2017) Global optimisation for dynamic systems using interval analysis. *Comput Chem Eng* 107:343–356

- Pitt JA, Gomoescu L, Pantelides CC, Chachuat B, Banga JR (2018) Critical assessment of parameter estimation methods in models of biological oscillators. *IFAC-PapersOnLine* 51:72–75
- Polisetty PK, Voit EO, Gatzke EP (2006) Identification of metabolic system parameters using global optimization methods. *Theor Biol Med Model* 3:1–15
- Raue A, Schilling M, Bachmann J, Matteson A, Schelke M, Kaschek D, Hug S, Kreutz C, Harms BD, Theis FJ, Klingmüller U, Timmer J (2013) Lessons learned from quantitative dynamical modeling in systems biology. *PLoS One* 8:e74335
- Rodriguez-Fernandez M, Kucherenko S, Pantelides C, Shah N (2007) Optimal experimental design based on global sensitivity analysis. *Comput Aided Chem Eng* 24:63–68
- Sass S, Mitsos A, Bongartz D, Bell IH, Nikolov NI, Tsoukalas A (2024) A branch-and-bound algorithm with growing datasets for large-scale parameter estimation. *European Journal of Operational Research*
- Schittkowski K (2002) *Numerical data fitting in dynamical systems*. Springer, Berlin
- Singer AB, Taylor JW, Barton PI, Green WH (2005) Global dynamic optimization for parameter estimation in chemical kinetics. *J Phys Chem A* 110:971–976
- Song Y, Khan KA (2022) Optimization-based convex relaxations for nonconvex parametric systems of ordinary differential equations. *Math Program* 196:521–565
- Sun J, Garibaldi JM, Hodgman C (2011) Parameter estimation using metaheuristics in systems biology: a comprehensive review. *IEEE/ACM Trans Comput Biol Bioinf* 9:185–202
- Tawarmalani M, Sahinidis N (2005) A polyhedral branch-and-cut approach to global optimization. *Math Program* 2:101–112
- Vassiliadis VS, Sargent RW, Pantelides CC (1994) Solution of a class of multistage dynamic optimization problems. 1. problems without path constraints. *Ind Eng Chem Res* 33:2111–2122
- Villaverde AF, Fröhlich F, Weindl D, Hasenauer J, Banga JR (2018) Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics* 35:830–838
- Villaverde AF, Henriques D, Smallbone K, Bongard S, Schmid J, Cicin-Sain D, Crombach A, Saez-Rodriguez J, Mauch K, Balsa-Canto E, Mendes P, Jaeger J, Banga JR (2015) BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. *BMC Systems Biology*, 9
- Wall L, Christiansen T, Orwant J (2000) *Programming Perl*. O'Reilly Media, Inc
- Wilhelm ME, Le AV, Stuber MD (2019) Global optimization of stiff dynamical systems. *AIChE J* 65:e16836
- Wodarz D, Nowak MA (1999) Specific therapy regimes could lead to long-term immunological control of hiv. *Proc Natl Acad Sci* 96:14464–14469
- Wächter A, Biegler LT (2006) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program* 106:25–57

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.