Systems biology

Benchmarks for identification of ordinary differential equations from time series data

Peter Gennemark[1,2,*] and Dag Wedelin[3]

[1]Department of Mathematical Sciences, University of Göteborg, SE-412 96 Göteborg, [2]Department of Mathematics, Uppsala University, SE-751 06 Uppsala and [3]Department of Computer Science and Engineering, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

ABSTRACT

Motivation: In recent years, the biological literature has seen a significant increase of reported methods for identifying both structure and parameters of ordinary differential equations (ODEs) from time series data. A natural way to evaluate the performance of such methods is to try them on a sufficient number of realistic test cases. However, weak practices in specifying identification problems and lack of commonly accepted benchmark problems makes it difficult to evaluate and compare different methods.

Results: To enable better evaluation and comparisons between different methods, we propose how to specify identification problems as optimization problems with a model space of allowed reactions (e.g. reaction kinetics like Michaelis–Menten or S-systems), ranges for the parameters, time series data and an error function. We also define a file format for such problems.

We then present a collection of more than 40 benchmark problems for ODE model identification of cellular systems. The collection includes realistic problems of different levels of difficulty w.r.t. size and quality of data. We consider both problems with simulated data from known systems, and problems with real data. Finally, we present results based on our identification algorithm for all benchmark problems. In comparison with publications on which we have based some of the benchmark problems, our approach allows all problems to be solved without the use of supercomputing.

Availability: The benchmark problems are available at www.odeidentification.org

Contact: peterg@chalmers.se

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

We consider the problem of identifying both the structure and the parameters of an ordinary differential equation (ODE) system from time series data. In recent years, there has been a significant increase in the number of reported methods approaching this problem in the biological literature (see Table 1 and the references).

For systems of realistic size and with realistic amount and quality of data this is a hard problem and heuristic algorithms are therefore

---

*To whom correspondence should be addressed.

typically employed. Such algorithms cannot guarantee that they give the best answer to a given problem. An established way to evaluate the performance of heuristic algorithms is to try them on a sufficient number of realistic test cases. However, lack of well defined and commonly accepted benchmark problems makes it difficult to evaluate and compare different methods:

- Problems are often incompletely and/or ambiguously specified, and personal communication is required to reconstruct essential aspects of the problems.

- Very few test problems are usually considered. This serves the purpose of demonstrating feasibility, but says little about general performance.

- Even when problems originate from the same system, different papers typically consider slight modifications of the original problem, which make direct comparisons difficult. An example is a test problem from Kikuchi *et al.*, 2003. Several recent publications (Cho *et al.*, 2006; Daisuke and Horton, 2006; Gennemark and Wedelin, 2007; Kimura *et al.*, 2005; Liu and Wang, 2008; Tsai and Wang, 2005; Tucker and Moulton, 2006) consider the same system, but usually with a more informative dataset.

To a considerable degree, the lack of benchmark problems is caused by a more fundamental underlying difficulty in how to specify such *identification problems* mathematically, and how to conveniently represent them in files. This is much more complex when both structure and parameters are unknown, compared with the well known case of parameter estimation in a given structure.

In particular, it is not sufficient to give only data from one or several experiments, or a *source model* from which such data can be simulated, see Figure 1A. In addition to the data, a model space must be defined, to specify the possible forms of the right-hand side of the ODEs and appropriate parameter ranges. Clearly, if we allow few possible reactions and narrow parameter ranges we obtain a simpler identification problem, and if we allow many possible reactions and wide parameter ranges we obtain a more difficult problem. Furthermore, known prior information about parts of the model may be given. Finally, we must define an objective function that also includes some notion of model complexity. This is needed since we are searching among different structures, and a simple maximum likelihood criterion is not sufficient and will

Table 1. Benchmark problems for ODE system identification

Problem name	Original problem reference	Source system/Model	#var	#exp	#pts	Noise (%)
simpleLin1		Simple linear system	3	3	13	0
simpleLin2				8	13	10
simpleFb1		Simple feedback loop (McKinney *et al.*, 2006)	3	4	7	0
simpleFb2				4	7	5
simpleFb3				1	7	0
simpleFb4	(McKinney *et al.*, 2006)			1	7	≈5[a]
osc1		An oscillator (Karnaukhov *et al.*, 2007)	3	1	41	0
osc2	(Karnaukhov *et al.*, 2007)			1	41	3
metabol1	(Gennemark and Wedelin, 2007)	A metabolic pathway (Arkin and Ross, 1995)	5	12	7	0
metabol2	(Gennemark and Wedelin, 2007)			12	21	10
metabol3	(Gennemark and Wedelin, 2007)			12	21	20
3genes1		A 3-step gene network (Moles *et al.*, 2003)	8	16	21	0
ss_cascade1	(Tsai and Wang, 2005)	A cascaded pathway (Voit, 2000)	3	8	41	0
ss_cascade2				4	41	0
ss_cascade3				8	41	5
ss_branch1	(Voit and Almeida, 2004)	A branched pathway (Voit, 2000)	4	3	21	0
ss_branch2	(Marino and Voit, 2006)			6	51	0
ss_branch3	(Tucker and Moulton, 2006)			5	20	0
ss_branch4	(Kutalik *et al.*, 2007)			4	20	0
ss_branch5	(Kutalik *et al.*, 2007)			4	20	2.5
ss_branch6	(Gonzalez *et al.*, 2007)			5	31	0
ss_5genes1	(Kikuchi *et al.*, 2003)	A genetic network (Hlavacek and Savageau, 1996)	5	10	11	0
ss_5genes2	(Gennemark and Wedelin, 2007)			10	9	20
ss_5genes3	(Gennemark and Wedelin, 2007)			10	3	0
ss_5genes4	(Kimura *et al.*, 2005)			15	11	0
ss_5genes5	(Daisuke and Horton, 2006)			10	11	0
ss_5genes6	(Cho *et al.*, 2006)			1	16	0
ss_5genes7	(Tucker and Moulton, 2006)			10	20	0
ss_5genes8	(Tsai and Wang, 2005) (Liu and Wang, 2008)			8	41	0
ss_15genes1		A genetic network (Maki *et al.*, 2001)	15	10	11	0
ss_15genes2				20	11	10
ss_30genes1		A genetic network (Maki *et al.*, 2001)	30	15	11	0
ss_30genes2	(Kimura *et al.*, 2005)			20	11	10
ss_30genes3	(Liu and Wang, 2008)			8	41	0
cytokine1	(McKinney *et al.*, 2006)	Immunologic data (Rock *et al.*, 2004)	4	1	7	10[a]
cytokine2				1	7	10[a]
ss_ethanolferm1	(Liu and Wang, 2008)	Ethanol fermentation (Wang *et al.*, 2001)	4	2	11–15	≈30[a]
ss_ethanolferm2				3	11–19	≈30[a]
ss_sosrepair1	(Cho *et al.*, 2006)	SOS repair system *Escherichia coli* (Ronen *et al.*, 2002)	6	1	50	10[a]
ss_sosrepair2				1	50	10[a]
ss_cadBA1	(Gonzalez *et al.*, 2007)	cadBA network in *E. coli* (Kuper and Jung, 2005)	5	1	25	<20[a]
ss_cadBA2				1	25	<20[a]
ss_clock1	(Daisuke and Horton, 2006)	Mice cell cycle (Barrett *et al.*, 2005)	7	1	12	≈10[a]
ss_clock2				1	12	≈10[a]

[a]Estimate from or assumption about data. See web site for further information.

#var, number of dependent variables; #exp, number of experimental conditions with different initial conditions and/or input functions; #pts, number of data-points per time series. Noise is added from a Gaussian distribution with SD given as a certain percentage (denoted Noise) of each experimental value. Problem names starting with 'ss_' correspond to S-systems. The last section lists problems with real data from biological systems.
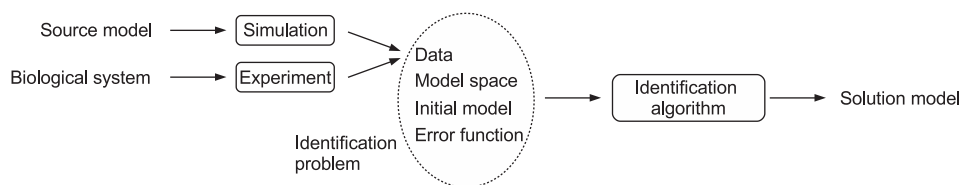
lead to over-fitting. Together, this defines the identification problem mathematically as an *optimization problem*, the *solution* of which is a model. An illustration of an identification problem is shown in Figure 1B, and the solution model is shown in Figure 1C.

Based on these considerations, we present a collection of more than 40 fully specified and publicly available benchmark problems for ODE system identification. The problems have mainly been collected and translated from recent publications, but we have also included new and carefully designed problems.

The collection includes both simple problems suitable for initial testing, and more complex problems with realistic size and quality of data. Most problems are based on simulated data from a known
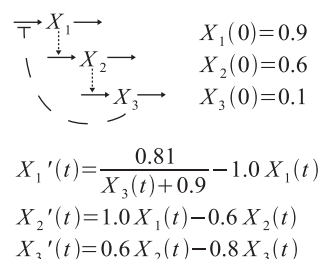
A ODE SYSTEM IDENTIFICATION

B AN IDENTIFICATION PROBLEM

C THE SOLUTION MODEL

Fig. 1. Identification of ODE systems. (**A**) An identification problem can be specified with real or simulated data from one or several experiments, a model space of allowed reactions occurring on the right-hand side of the ODEs, an initial model representing prior knowledge and an error function. (**B**) An example of an identification problem with a model space of three traditional chemical reaction types, an error function where L = likelihood function, λK = structural complexity term (λ = constant and K = number of model parameters) and an initial model with no prior information. (**C**) An example of a solution model.

source model, which is important for evaluation purposes. However, several problems are also based on real data.

The collection consists of problems of two different kinds. The first kind are problems whose solutions are models based on traditional chemical rate equations like unimolecular mass action and Michaelis–Menten's kinetics. An example of such a system with three variables is shown in Figure 1C. Typically for systems of this kind, the rate equations are composed as linear combinations of known reaction types that can be non-linear in both the parameters and in the state variables.

The other kind of problem is where the solutions are defined on a fixed form, like S-systems which is currently the most popular form for identification. The S-system formalism (Savageau, 1976; Voit, 2000) is based on approximating kinetic laws with multivariate power-law functions. A model consists of n non-linear ODEs and the generic form of equation i reads

$$X_i'(t) = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}}(t) - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}}(t) \qquad (1)$$

where X is a vector (length n) of dependent variables, α and β are vectors (length n) of non-negative rate constants and g and h are matrices ($n \times n$) of kinetic orders, that can be negative as well as positive. The structure of the system is defined by the non-zero elements of g and h, and for realistic models most of these parameters are zero.

To our knowledge, this is the first article to address how to best represent this kind of identification problems as optimization problems and to define a suitable file format. As far as we know, this is also the first collection of benchmark problems for ODE system identification, where both structure and parameters are unknown.

Another paper that recognizes the importance of systematic evaluation of identification algorithms, is Mendes *et al.* (2003),

which presents a system for generating realistic artificial gene networks and simulated data, and a large number of problems generated by this system. In comparison, our benchmark problems are typically smaller and address the more specific need of evaluating algorithms identifying both structure and parameters, fully specifying problems as optimization problems. For parameter estimation only, several benchmark problems are publicly available, e.g. in the data base EASY-FIT (Schittkowski, 2002).

The two kinds of systems we consider, chemical rate equations and S-systems, have previously been used for identification of ODE models as indicated by references in Table 1. In addition to these kinds of systems, other fixed forms have been considered in the literature. One example is Generalized Mass Action (GMA) models (Tucker *et al.*, 2007; Voit, 2000), and another is linear dynamical models with non-linear transfer functions of sigmoid type (D'haeseleer, 2000; Nelander *et al.*, 2008; Wahde and Hertz, 2000). However, at this stage, we have chosen to focus on the more frequently used S-systems as one representative fixed form system.

We finally note that identification of ODEs where more basic operands and operators are used as structural components have been considered (Bongard and Lipson, 2007).

2 SPECIFYING IDENTIFICATION PROBLEMS AS OPTIMIZATION PROBLEMS

We now outline how we define identification problems mathematically. Our first and most important concern has been that an identification problem must be defined unambiguously as an optimization problem. This ensures that a problem has a well defined solution independently of any method for solving it, and that the task of *modelling* identification problems is separated from the algorithmic task of *solving* given problems. It is to be noted

that if we fail to explicitly specify the entire problem, missing parts have to be supplied by the user or be implicitly defined by the algorithm, resulting in confusion and lack of reproducibility.

Our second concern has been to find a reasonably simple standard form that can still represent a wide range of identification problems. We have then chosen the following way to structure and specify our problems.

- *Data*. Time series data for one or several experiments. In the case of several experiments, they may differ with respect to initial values of the variables and/or input functions. For each data-point, the standard deviation is also given in the problem specification.

- *Model space*. The model space determines the allowed form of the right-hand side of the ODEs.
  For models based on traditional chemical rate equations, each ODE in the model is assumed to be a sum of a number of reactions. The possible reactions must belong to a subset of predefined reaction types, where each allowed reaction type is specified by its name, a subset of possible input variables and ranges of allowed parameter values. The allowed reaction types can be specified individually for each state variable. As an example, in Figure 1B, the allowed reactions types are a unimolecular mass action reaction, a Michaelis–Menten reaction, and a simplified Hill equation. For reactions having multiple input variables, e.g. a bimolecular mass action reaction with equation $k_1 X_i X_j$, it is implicitly assumed that i and j are not equal (to consider equality as in problem osc1 and osc2, we define an additional reaction type $k_1 X_i^2$).
  The model space of an S-system is simply defined by lower and upper bounds for each element in the parameter vectors (α and β) and matrices (g and h). Sometimes, additional constraints are required. For example, three of the benchmark problems include an additional constraint of type $\{g_{i,j} \in [-3, 3], g_{i,j} \neq 0\}$ (there is an interaction between variable i and j but the direction is unknown).
  Finally, we also define lower and upper bounds for the initial data-point in each time series. For noisy data, these bounds were set to ± 2 SDs. Hence, for noisy data there is one additional parameter for each time series, but these parameters are typically bound tighter compared with the model parameters.

- *Initial model*. It is convenient to allow definition of an initial model, corresponding to prior knowledge of the system. The initial model is described as known reactions (terms) on the right-hand side of the ODEs. Also reactions from outside the model space can be included in the initial model.
  In principle, one can also think of prior information in the form of starting points for iterative algorithms and thus not technically a part of the defined problem. No such information is assumed known in our current problems.

- *Error function*. We have chosen to minimize

$$-L(\hat{X}|k) + \lambda K. \tag{2}$$

The first term is the negative log-likelihood of the experimental data, and the second term is a term that penalizes structural complexity of the model. This kind of error function is common, and is related to several different proposed methods for handling model complexity (Crampin *et al.*, 2004).

In detail, L is the log-likelihood, $\hat{X}$ denotes the experimental data, k is a vector of parameters, λ is a constant and K is the number of parameters. By assuming independent and normally distributed measurement errors and disregarding constant terms we can express the log-likelihood for one time series as

$$L(\hat{X}_j|k) = -\frac{1}{2} \sum_i \left(\frac{X_j(t_i) - \hat{X}_j(t_i)}{\sigma_j(t_i)} \right)^2 \tag{3}$$

where i indexes the measurement points, and where X_j, $\hat{X}_j$ and σ_j denotes simulated data, experimental data and SD for variable j, respectively. The total log-likelihood $L(\hat{X}|k)$ is defined by summing over all variables and all experiments. For models based on chemical rate equations, K is simply the total number of parameters on the right-hand side of the ODEs. For S-systems, it is natural to define K as the total number of non-zero elements in g and h plus the number of parameters in α and β.

To establish the relationship with standard optimization terminology, the model space and and the initial model define the *feasible set*, and the data together with the error function define the *objective function*.

2.1 File format for identification problems

In order to work with identification problems and to provide them as input to identification algorithms, we also need to represent such problems in files, and it is highly desirable that an entire problem can be represented in a single file. Since an identification problem is an optimization problem and not a model, common model formats such as SBML are not applicable, and no existing format known to us can be used for this purpose. However, the output of an identification algorithm is a model and can be represented in SBML. Also, if a partially known initial model is available in SBML, an identification algorithm can input that part in SBML and the rest of the identification problem in our format.

The file format we have designed for the identification problems can be seen as a special case of a more general format that is currently developed as a separate project. The format attempts to be self documenting and easy to read both by the human and the computer. Compared with a typical XML-equivalent, more structure is explicit and it is easier to read. It bears some resemblance to how data structures are built in a computer programme, but without the explicit use of pointers. An extract of the format is given below:

```
{
// VARIABLES
variable_1 has name = x1 is inputVariable
variable_2 has name = x2 is inputVariable
variable_3 has name = x3 is dependent
...
// MODEL SPACE OF VARIABLE 3
possibleReaction_3 of variable_3
has type = biMolecularMassAction
has spaceOfVariable X1 = memberOfSet_2
has spaceOfVariable X2 = memberOfSet_1
has rangeOfParameter k1 = range_1
...
```

```
// EXPERIMENT 1
sample_1 of experiment_1
has time =  0.00
has variable_ =   3.00  2.00  ...
has sdev of variable_ =  0.00 0.00 ...
...}
```

This particular extract describes first some variables and then a possible reaction in the model space. Finally, a sample from the first experiment is given.

The format can also be used to specify a parameter estimation problem only. Method specific parameters (like random seed) can be defined in the same file or separately. Finally, we can easily extend the format to describe new classes of problems, for example with compartment modelling and system modifications like gene deletions.

We anticipate that the exact mathematical form of the identification problems, as well as the file format, will be extended over time. Up-to-date information and detailed documentation is therefore available on the web site, as well as a simple parser for the file format.

3   THE COLLECTION OF BENCHMARK PROBLEMS

An overview of the collection of benchmark problems is presented in Table 1. The problems have different sources, and in most cases several different problems are given for each system.

The number of variables in combination with the amount of data and noise level, give a rough indication of the difficulty of each problem. The first problem for each system generally has a lower level of difficulty than the subsequent problems, and is hence a good starting point when evaluating new methods.

A part of the collection is based on identification problems considered in the literature, which is dominated by problems based on S-systems, probably because these are relatively easy to represent. By including translations of these problems that were previously not easily accessible, the collection has an immediate value for those actively developing algorithms in this area. However, we balance the S-system dominance by also defining new problems based on traditional chemical reactions that are more frequently used in the biological modelling community. We note that research in this area is in an early stage, and in our experience identification algorithms for S-systems can relatively easily be modified for other kinds of models. Our proposed way of specifying identification problems can then facilitate and contribute to a more widespread use of problems based on other kinds of models.

For problems that have been collected and translated from another publication, we have taken care to match the original problem as closely as available information and the standardization to our formalism allow, see the Supplementary Material for details. We have replaced some non-standard objective functions with our simple function, and relaxed a few unusual constraints. Noisy data have been simulated from random distributions, and in one case we have also reduced an excessive and unrealistic number of data-points in the original problem. However, the solutions in the next section confirm that the changes have not negatively affected the quality of the solutions when compared directly with the source models. We note that our problems never have more data or a smaller model space, so they are in this sense at least as difficult as the original problems.

3.1   Best known solutions

In Table 2 we report the best known solutions for all benchmark problems, and we will now describe the detailed information in the table. The solutions have been found using our own identification algorithm (Gennemark and Wedelin, 2007). All solution models as well as source models are available in SBML on our web site.

For problems with noisy or otherwise insufficient data, an optimal solution has slightly different parameter values and sometimes also different structure than the source model. To clarify this for our solutions, Table 2 gives the error of the source model from which data was simulated. We can see that for all problems based on simulated data the found solutions have approximately equal or lower error than that of the source model. We also give the number of FP and FN interactions, to indicate if the found structure is the same as the source model. If the structure is the same as for the source model, we also give the largest deviation (LD) in percent for any parameter compared with the source model. For the problems based on real data, no source model is available, and no such comparisons can be made.

For our algorithm, as well as for most algorithms in the literature, the results may differ between runs depending on a random seed. With our approach, the main difference between runs is that sometimes the best structure is not found due to the heuristic nature of the search algorithm. However, for two solutions with the same structure, the difference in the parameters is negligible for all tested problems. To indicate the stability of the algorithm, we therefore report the proportion of runs where the best structure was found. Note, however, the fact that solutions are stable does not generally imply that they are optimal.

Computation times are reported as the average computation time for runs with different random seeds, scaled to a 1 GHz processor (the actual runs were performed on a Pentium IV, 2.13 GHz). The times can be improved by adjusting the method parameters for individual problems, but we report the computation time using our standard settings, see the Supplementary Material.

For the benchmark problems that have been adapted from problems in the literature, the final columns report available results for these original problems.

In addition to the numerical details, Table 2 allows us to draw the following qualitative conclusions about the benchmark problems.

- Every problem with data from a known source model is well formulated in the sense that the solution model is similar to the source model, in terms of the FP/FN/LD values.

- Every solution model is at least as similar to the source model as any previously reported solution. This confirms that the translation to our standard format works well, and also that our solutions are indeed the best known solutions to the benchmark problems.

- Every problem can be solved within hours on an ordinary computer with our approach, so there is no need for supercomputing. This was not previously established since some previous results for the original problems required running times of several months if run on a single processor.

Table 2. Best known solutions to the benchmark problems

Problem	Error of source model $-L+\lambda K$	Best known solution[a]					Solution of original problem (if available)	
		Error $-L+\lambda K$	Residual $-L$	Structure FP/FN/LD	Stability between runs	Average time (s) (1 GHz)	Structure FP/FN/LD	Time (s) (1 GHz)
simpleLin1	8	8	0	0/0/0	5/5	17		
simpleLin2	187.4	140.2	132.2	0/0/–	5/5	86		
simpleFb1	7	7	0	0/0/0	5/5	14		
simpleFb2	45.46	42.14	34.14	1/1/–	3/5	36		
simpleFb3	7	7	0	0/0/0	3/5	41		
simpleFb4	18.39	7.071	0.0713	0/0/–	2/5	47	0/0/–	–
osc1	6	6.149	0.1491	0/0/0	5/5	18		
osc2	122.1	87.34	57.34	0/0/0	4/5	24	–	–
metabol1	30	30	0	0/0/0	3/3	5400		
metabol2	1214	751.6	601.6	0/0/–	2/3	11 000		
metabol3	1442	861.1	696.1	2/1/–	1/5[b]	13 000		
3genes1	39	39	0	0/0/0	3/3	35 000		
ss_cascade1	14	14	0	0/0/0	5/5	180	0/0/300	8900
ss_cascade2	14	14	0	0/0/0	4/5	130		
ss_cascade3	498.9	476.7	462.7	1/1/–	1/5[b]	860		
ss_branch1	17	17	0	0/0/0	5/5	68	0/0/15	2200
ss_branch2	17	17	0	0/0/0	5/5	110	0/0/5	35 000
ss_branch3	17	17	0	0/0/0	5/5	150	0/0/2	15 000
ss_branch4	17	17	0	0/0/0	5/5	71	0/0/0	25 000
ss_branch5	211.3	142.1	122.1	4/1/–	3/5	300	–	–
ss_branch6	18	18	0	0/0/0	5/5	90	0/0/5	1100
ss_5genes1	23	23	0	0/0/0	5/5	600	1/0/–	2.4E+8
ss_5genes2	300.3	211.0	183.0	5/0/–	2/5[b]	2100		
ss_5genes3	23	23	0	0/0/0	5/5	380		
ss_5genes4	23	23	0	0/0/0	5/5	640	37/0/–	40 000
ss_5genes5	23	23.00	1.14E−3	0/0/0.02	5/5	400	4/2/–	7200
ss_5genes6	23	23	0	0/0/0	5/5	130	0/0/50	56 000
ss_5genes7	23	23	0	0/0/0	5/5	190	0/0/0.2	15 000
ss_5genes8	28	28	0	0/0/0	5/5	440	0/0/5	26 000
							0/0/2.5	7000
ss_15genes1	62	62	0	0/0/0	5/5	1500		
ss_15genes2	2010	1783	1478	3/4/–	2/5[b]	18 000		
ss_30genes1	128	128	0	0/0/0	5/5	7900		
ss_30genes2	4098	3628	2993	6/7/–	1/3[b]	94 000	242/10/–	6.2E+6
ss_30genes3	128	128	0	0/0/0	3/3	18 000	0/0/25	8.6E+6
cytokine1		25.92	17.92		1/5[c]	11	–	–
cytokine2		42.17	32.17		1/5[c]	23		
ss_ethanolferm1		127.4	110.4		1/5[c]	190	err>127.4[d]	–
ss_ethanolferm2		1308	1292		1/5[c]	360		
ss_sosrepair1		2642	2611		1/5[c]	510	err>2642[d]	2.7E+5
ss_sosrepair2		2823	2789		1/5[c]	470		
ss_cadBA1		750.6	726.6		1/5[c]	250	err>750.6[d]	>540
ss_cadBA2		709.1	687.1		1/5[c]	260		
ss_clock1		928.4	803.4		1/5[c]	360	–	15 000
ss_clock2		814.5	649.5		1/5[c]	440		

[a]Best known solutions, based on comparsion to the solutions of the original problems. For ss_30genes2, the similarity with the source model strongly indicates that our solution is the best, but since we do not have access to the solution of the original problem this is not confirmed.

[b] The majority of runs have an error below the error of the source model, but differ slightly in structure.

[c]Real datasets with relatively few experiments and/or data-points making many similar models reasonable, increasing the variability in found solutions.

[d]Since no source model is available, we have evaluated the original solution with our error function, and it has an error higher than our solution.

The error of the source model refers to the error of the model from which data was simulated. The best result is taken from several runs with various random seeds. The error, the negative log-likelihood and the number of false positives and negative (FP/FN) reactions are reported. Stability measures the frequency of runs for which the best structure is obtained. Computation time in seconds is given as the average of several runs.

4 CONCLUSIONS

We have proposed a way to unambiguously specify ODE identification problems as mathematical optimization problems, and have also defined a file format making it easy to document and exchange such problems.

The benchmark problems include translations of problems that have been frequently used in the development of algorithms, as well as new problems based on chemical rate equations. The problems, their best solutions and the source models from which problem data was simulated are all available on our web site. This makes it easy for everyone to evaluate and compare algorithms in a reproducible way. To maintain a useful collection of problems of various types, we intend to extend the collection, and are interested in contributions and feedback from others, see the web site.

Finally, our own solutions show that all the benchmark problems can be solved in reasonable time, in contrast to results for some of the original problems where supercomputing was required.

REFERENCES

Arkin,A.P. and Ross,J. (1995) Statistical construction of chemical reaction mechanisms from measured time-series. *J. Phys. Chem.*, **99**, 970–979.

Barrett,T. *et al*. (2005) NCBI GEO: mining millions of expression profiles–database and tools. *Nucleic Acids Res.*, **33**, D562-D566.

Bongard,J. and Lipson,H. (2007) Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA*, **104**, 9943–9948.

Cho,D.Y. *et al*. (2006) Identification of biochemical networks by S-tree based genetic programming. *Bioinformatics*, **22**, 1631–16340.

Chou,I.C. *et al*. (2006) Parameter estimation in biochemical systems models with alternating regression. *Theor. Biol. Med. Model*, **3**, 25.

Crampin,E.J. *et al*. (2004) Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Prog. Biophy. Mol. Biol.*, **86**, 77–112.

Daisuke,T. and Horton,P. (2006) Inference of scale-free networks from gene expression time series. *J. Bioinform. Comput. Biol.*, **4**, 503–514.

D'haeseleer,P. (2000) *Reconstructing Gene Networks from Large Scale Gene Expression Data*. PhD Thesis. University of New Mexico, Albuquerque, NM, USA.

Gennemark,P. and Wedelin,D. (2007) Efficient algorithms for ordinary differential equation model identification of biological systems. *IET Syst. Biol.*, **1**, 120–129.

Gonzalez,O.R. *et al*. (2007) Parameter estimation using Simulated Annealing for S-system models of biochemical networks. *Bioinformatics*, **23**, 480–486.

Hlavacek,W.S. and Savageau,M.A. (1996) Rules for coupled expression of regulator and effector genes in inducible circuits. *J. Mol. Biol.*, **255**, 121–139.

Karnaukhov,A.V. *et al*. (2007) Numerical Matrices Method for nonlinear system identification and description of dynamics of biochemical reaction networks. *Biophys. J.*, **92**, 3459–3473.

Kikuchi,S. *et al*. (2003) Dynamic modeling of genetic networks using genetic algorithm and S-system, *Bioinformatics*, **19**, 643–650.

Kimura,S. *et al*. (2005) Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*, **21**, 1154–1163.

Kuper,C. and Jung,K. (2005) CadC-mediated activation of the cadBA promoter in Escherichia coli. *J. Mol. Microbiol. Biotechnol.*, **10**, 26–39.

Kutalik,Z. *et al*. (2007) S-system parameter estimation for noisy metabolic profiles using newton-flow analysis. *IET Syst. Biol.*, **1**, 174–180.

Liu,P.K. and Wang,F.S. (2008) Inference of biochemical network models in S-system using multiobjective optimization approach. *Bioinformatics*, **24**, 1085–1092.

Maki,Y. *et al*. (2001) Development of a system for the inference of large scale genetic networks. *Pac. Symp. Biocomput.*, 446–458.

Marino,S. and Voit,E.O. (2006) An automated procedure for the extraction of metabolic network information from time series data. *J. Bioinform. Comput. Biol.*, **4**, 665–691.

McKinney,B.A. *et al*. (2006) Hybrid grammar-based approach to nonlinear dynamical system identification from biological time series. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, **73**, 021912.

Mendes,P. *et al*. (2003) Artificial gene networks for objective comparison of analysis algorithms. *Bioinformatics*, **19** (Suppl. 2), ii122–ii129.

Moles,C.G. *et al*. (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.*, **13**, 2467–2474.

Nelander,S. *et al*. (2008) Models from experiments: combinatorial drug perturbations of cancer cells. *Mol. Syst. Biol.*, **4**, 216.

Rock,M.T. *et al*. (2004) Adverse events after smallpox immunizations are associated with alterations in systemic cytokine levels. *J. Infect Dis.*, **189**, 8.

Ronen,M. *et al*. (2002) Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl Acad. Sci. USA.*, **99**, 10555–10560.

Savageau,M.A. (1976) *Biochemical systems analysis: a study of function and design in molecular biology*. Addison-Wesley, Reading, MA.

Schittkowski,K. (2002) Numerical data fitting in dynamical systems: a practical introduction with applications and software. In *Applied Optimization*. Donald,H. and Pardalos,P.M. (eds.), vol. **77**. Kluwer Academic, Dordrecht.

Tsai,K.Y. and Wang,F.S. (2005) Evolutionary optimization with data collocation for reverse engineering of biological networks. *Bioinformatics*, **21**, 1180–1188.

Tucker,W. and Moulton,V. (2006) Parameter reconstruction for biochemical networks using interval analysis, *Reliable Computing*, **12**, 389–402.

Tucker,W. *et al*. (2007) Estimating parameters for generalized mass action models using constraint propagation. *Math. Bioscience*, **208**, 607–620.

Voit,E.O. (2000) Parameter estimation. *Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press, Cambridge, pp. 176–184.

Voit,E.O. and Almeida,J. (2004) Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics*, **20**, 1670–1681.

Wahde,M. and Hertz,J. (2000) Coarse-grained reverse engineering of genetic regulatory networks. *Biosystems*, **55**, 129–136.

Wang,F.S. *et al*. (2001) Hybrid differential evolution for problems of kinetic parameter estimation and dynamic optimization of an ethanol fermentation process. *Ind. Eng. Chem. Res.*, **40**, 2876–2885.