

A benchmark problem for methods in reverse engineering and model discrimination

A. Kremling*
Max-Planck-Institut for
Dynamics of Complex
Technical Systems
Sandtorstr. 1
39106 Magdeburg, Germany
kre@mpi-
magdeburg.mpg.de

T. Sauter
Institute for System Dynamics
and Control Engineering
University of Stuttgart,
Stuttgart,
Germany
sauter@isr.uni-
stuttgart.de

E. Bullinger
Institute for Systems Theory in
Engineering
University of Stuttgart,
Stuttgart,
Germany
eric@ist.uni-stuttgart.de

ABSTRACT

A benchmark problem is described for the reconstruction of biochemical networks given sampled experimental data. The growth of the organisms is described in a bio-reactor where one substrate is feed into the reactor with a given feed rate and feed concentration. Measurements for some intracellular components are provided representing a small biochemical network. Problems of model discrimination, parameter estimation, and identifiability are addressed. Feed rate and feed concentration could be used as control inputs. To verify calculated input profiles an interactive web-site (<http://www.sysbio.de/projects/benchmark/>) is provided.

1. INTRODUCTION

The analysis of metabolic and regulatory pathways with mathematical models contributes to the better understanding of the behavior of metabolic processes [2]. The setup of the structure of the model, i.e. the stoichiometry of the biochemical reaction network, is mainly based on data from database systems or from literature. Recent efforts in measurement technologies like cDNA array data or 2gel electrophoresis [1] will enable researchers to produce time courses of a number of substances from inside the cell. Having available such data, a challenging task is to identify the structure of the network (“reverse engineering”) and - if two or more model structures are suited to describe the experimental data - to design new experiments which will allow to discriminate between the model variants. Further problems comprise identifiability of the model parameters, sensitivity of the parameters and metabolic design [3]. Here, we present *in silico* experimental data for an organism growing in a chemostat (Figure 1). After reaching a steady state, the dilution rate q_{in} , q_{out} as well as the concentration of the substrate in the feed c_{in} are changed. Measurements are available for three metabolites $M1$, $M2$ and $M3$ representing a small biochemical network of the organism. The network structure has to be identified on the basis of these data and a new experiment has to be designed to select to most feasible model structure. Measurement for biomass X and substrate S are also available.

*Corresponding author

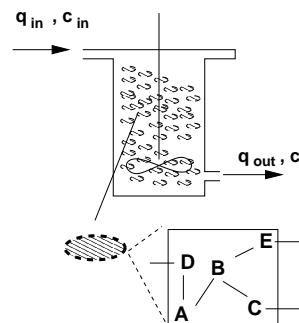


Figure 1: Scheme of the bio-reactor. Inputs are feed rates q_{in} , q_{out} and feed concentration c_{in} . Biomass is assumed to homogeneously distributed in the reactor. A biochemical reaction network is responsible for growth and cell division.

2. BENCHMARK PROBLEM DEFINITIONS

2.1 Initial conditions

Figure 2 shows the time course of the metabolites $M1$, $M2$, and $M3$ as well as the time course of biomass X and substrate S . The conditions during the chemostat experiment are summarized in Table 1. The molecular weight for the substrate used is 342.3 mol/g. The initial conditions for biomass and substrate are 0.1 g/l and 2.0 g/l, respectively. The volume of the bio-reactor was held constant at 1.0 l for the given time series (the maximal working volume of the reactor is $V_{max} = 5$ l).

2.2 Measurement

Measurement are sampled every 2 hours. To allow realistically complex behavior, the following procedure was employed. A set of kinetic parameters was chosen for the (hidden) network. After a simulation run with the starting conditions from above the measured data were generated: With a random number $rand$ the absolute value was modified according to $\hat{x} = x(1 + rand)$, $rand$ is normally distributed with mean value $\bar{m} = 0$, and standard deviation $\sigma = 0.1$.

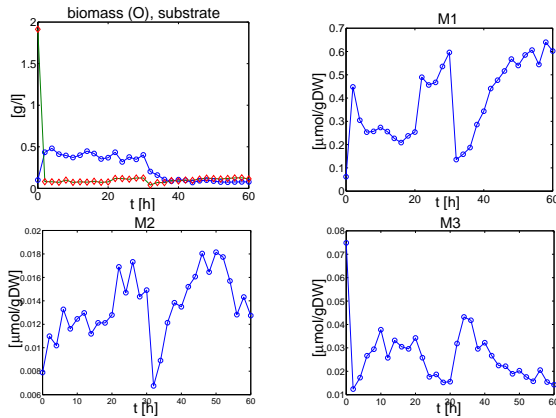


Figure 2: Time series data for biomass and substrate (upper left), for substance M1 (upper right), for substance M2 (lower left) and for substance M3 (lower right). Numerical values of the data are given in the appendix.

Table 1: Conditions during continuous culture experiment (the volume of the bio-reactor was held constant at 1.0l).

time	input	time	input
0 - 20 h	$q_{in} = 0.25$ l/h $q_{out} = 0.25$ l/h $c_{in} = 2.0$ g/l	20 - 30 h	$q_{in} = 0.35$ l/h $q_{out} = 0.35$ l/h $c_{in} = 2.0$ g/l
30 - 60 h	$q_{in} = 0.35$ l/h $q_{out} = 0.35$ l/h $c_{in} = 0.50$ g/l		

3. BENCHMARK PROBLEM APPLICATIONS

3.1 Network identification

Based on the measurement of substances M1, M2, and M3 the structure of network has to be identified. Note that all substances are able to reach a steady-state. The rate of synthesis or degradation of the substances may be constant or may depend on the metabolites or the extracellular substrate.

3.2 Model discrimination

If some more information on the network is existing another problem in reverse engineering can be addressed. The following biochemical facts can be used for further studies:

- Metabolite M1 is the first substance after uptake. The transport mechanism was identified as a Michaelis-Menten reaction law with parameters given in Table 2.
- Substance M3 acts as an enzyme E converting metabolite M1 in M2. The reaction is irreversible and the affinity (dissociation constant) of M1 was determined (see also Table 2).
- Degradation of M2 is also identified as a Michaelis-Menten reaction law with parameters given in Table 2.

It is assumed that flux from M2 is responsible for the entire biomass:



- The enzyme is subject to control (control of activity or control of synthesis).

Using two different (or more) models describing the control of the enzyme, the design of a new experiment is required to select the most feasible model structure.

Table 2: Kinetic parameters for synthesis of M1, degradation of M2 and the affinity of M1 to enzyme E.

synthesis of	M1
values:	$r_{max} = 24000$ μmol/gDW h $K = 0.4437$ μmol/gDW
affinity	M1 - E
value:	$K = 12.2$ μmol/gDW
degradation of	M2
values:	$r_{max} = 3 \cdot 10^6$ μmol/gDW h $K = 10.0$ μmol/gDW

3.3 Problem extensions

3.3.1 Further control inputs

To design a new experiment control variables u have to be fixed. To direct the chemostat, the feed rates q_{in} , q_{out} as well as the feed concentration c_{in} are possible variables.

For some strategies it might be necessary to include a control variable that directly influences the enzyme either by alteration of the activity or the rate of synthesis. The following scenario allows to control enzyme synthesis: if an enzyme is under control, at least one transcription factor is necessary to activate or to repress gene expression. The amount of the transcription factor can be varied if we assume that the expression of the transcription factor is under full control, e.g. by given an inducer in distinct concentration.

3.3.2 Metabolic design

Another extension would be to expand the proposed biochemical network through the addition of new pathways. Such a metabolic redesign would allow the redirection of fluxes to achieve a desired impact.

3.3.3 Identifiability

To design a new experiment inputs and outputs must be chosen in such a way that parameters could be identified. Furthermore parameters could only be estimated with high quality if the control inputs direct them into sensitive regions.

3.3.4 Modeling approach

Although the problem is suited for models with ordinary differential equations other approaches using e.g. automaton may also lead to interesting results.

3.3.5 Education

The problem could be used as a study in metabolic modeling to illustrate methods in model set-up, model analysis and system design.

4. INTERACTIVE WEB-SITE

To verify calculated input profiles an interactive web-site is provided under <http://www.sysbio.de/projects/benchmark/>. The site offers the possibility to enter a vector of time points and corresponding values for the input profiles for q_{in} , q_{out} , and c_{in} as well as sampling time points (in h). Initial conditions for all states must be given also. Outputs are the time vector at the given sampling time points and a vector of all states with random noise. The time series data are shown in a number of plots and can be download also.

5. ADDITIONAL AUTHORS

S. Fischer, email address: fischer@mpi-magdeburg.mpg.de,
 F. J. Doyle, email address: doyle@engineering.ucsb.edu,
 F. Allgöwer, email: allgower@ist.uni-stuttgart.de and
 E.D. Gilles, email: gilles@mpi-magdeburg.mpg.de.

6. REFERENCES

- [1] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292:929–9334, 2001.
- [2] H. Kitano. Perspectives on systems biology. *New Generation Computing*, 18(3):199–216, 2000.
- [3] J. Stelling, A. Kremling, M. Ginkel, K. Bettenbrock, and E. D. Gilles. Towards a virtual biological laboratory. In H. Kitano, editor, *Foundations of Systems Biology*, chapter 9, pages 189–212. The MIT Press, 2001.

APPENDIX

A. MEASUREMENT

time [h]	X [g/l]	S [g/l]
0	0.1008	1.9134
2.0000	0.4345	0.0805
4.0000	0.4811	0.0791
6.0000	0.4114	0.0734
8.0000	0.3956	0.0990
10.0000	0.3714	0.0724
12.0000	0.3995	0.0782
14.0000	0.4477	0.0752
16.0000	0.4190	0.0853
18.0000	0.3540	0.0725
20.0000	0.3690	0.0781
22.0000	0.4345	0.1195
24.0000	0.3183	0.1178
26.0000	0.3767	0.1099
28.0000	0.3489	0.1243
30.0000	0.4019	0.1249
32.0000	0.2023	0.0403
34.0000	0.1595	0.0703

36.0000	0.1068	0.0691
38.0000	0.0868	0.0933
40.0000	0.1047	0.0893
42.0000	0.0967	0.1000
44.0000	0.0714	0.0965
46.0000	0.0916	0.1122
48.0000	0.0992	0.1234
50.0000	0.0877	0.1180
52.0000	0.0766	0.1133
54.0000	0.0747	0.1196
56.0000	0.0769	0.1256
58.0000	0.0786	0.1269
60.0000	0.0781	0.1138

M1 M2 M3

0.0620	0.0079	0.0749
0.4479	0.0110	0.0124
0.3045	0.0102	0.0173
0.2534	0.0133	0.0266
0.2569	0.0116	0.0294
0.2736	0.0125	0.0378
0.2561	0.0130	0.0257
0.2268	0.0112	0.0332
0.2086	0.0121	0.0305
0.2375	0.0121	0.0296
0.2539	0.0128	0.0342
0.4895	0.0169	0.0258
0.4561	0.0147	0.0176
0.4673	0.0173	0.0187
0.5358	0.0144	0.0152
0.5961	0.0149	0.0156
0.1357	0.0067	0.0319
0.1584	0.0089	0.0432
0.1873	0.0121	0.0418
0.2860	0.0138	0.0296
0.3434	0.0135	0.0322
0.4408	0.0152	0.0267
0.4767	0.0161	0.0225
0.5163	0.0180	0.0222
0.5675	0.0165	0.0189
0.5399	0.0181	0.0202
0.5851	0.0177	0.0176
0.6062	0.0157	0.0157
0.5443	0.0128	0.0205
0.6399	0.0143	0.0154
0.6020	0.0127	0.0142

Values of $M1$, $M2$ and $M3$ are in [$\mu\text{mol/gDW}$]. A file with the presented data can be download from the above mentioned web-site.