Experiment Selection for the Discrimination of Semi-Quantitative Models of Dynamical Systems

Ivayla Vatcheva, Olivier Bernard, Hidde de Jong, Nicolaas Mars

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Experiment Selection for the Discrimination of Semi-Quantitative Models of Dynamical Systems

Ivayla Vatcheva — Olivier Bernard — Hidde de Jong — Nicolaas Mars

N° 4940

Septembre 2003

THÈMES 3 et 4

ISSN 0249-6399 ISRN INRIA/RR--4940--FR+ENG

INRIA
RHÔNE-ALPES

Experiment Selection for the Discrimination of Semi-Quantitative Models of Dynamical Systems

Ivayla Vatcheva* , Olivier Bernard , Hidde de Jong , Nicolaas Mars†

Thèmes 3 et 4 — Interaction homme-machine,
images, données, connaissances — Simulation et optimisation
de systèmes complexes
Projets HELIX,COMORE

Rapport de recherche n° 4940 — Septembre 2003 — 40 pages

Abstract: Modeling an experimental system often results in a number of alternative models that are all justified by the available experimental data. In order to discriminate between these models, additional experiments are needed. We present a method for experiment selection that helps in discriminating between differential equation models of experimental systems in a systematic and efficient way. The method generalizes upon previous work on model discrimination in that it deals with semi-quantitative differential equations, which use interval bounds on parameter values and envelopes for functional relations. The model discrimination method is based on an entropy criterion for the selection of the most informative experiment. The applicability of the method to real-world problems is illustrated by means of an example in population biology, the discrimination of competing models of the growth of phytoplankton in a bioreactor.

Key-words: semi-quantitative modeling and simulation, model discrimination, information theory, population biology

* LION Bioscience AG, Heidelberg, Germany
† Materials Science Centre, Department of Mathematics and Natural Sciences, Rijksuniversiteit Groningen, Groningen, the Netherlands

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 —Télécopie +33 4 76 61 52 52

Sélection d'expériences pour la discrimination de modèles semi-quantitatifs de systèmes dynamiques

Résumé : La modélisation des systèmes expérimentaux donne souvent lieu à plusieurs modèles alternatifs, tous justifiés par les données expérimentales disponibles. Afin de pouvoir discriminer ces modèles entre eux, des expériences supplémentaires sont nécessaires. Nous présentons une méthode pour la sélection d'expériences qui aide, de façon systématique et efficace, à discriminer les modèles à équations différentielles d'un système expérimental. Par sa capacité à traiter des équations différentielles semi-quantitatives en utilisant des intervalles pour les valeurs de paramètres et des enveloppes pour des relations fonctionnelles, la méthode est une généralisation des travaux précédents sur la discrimination de modèles. La méthode de discrimination de modèles est basée sur un critère d'entropie pour la sélection de l'expérience la plus informative. L'applicabilité de la méthode à des problèmes réels est illustrée par un exemple en biologie des populations : la discrimination de modèles concurrents de croissance de phytoplancton en bioréacteur.

Mots-clés : modélisation et simulation semi-quantitative, discrimination de modèles, théorie de l'information, biologie des population

1 Introduction

The formulation of an adequate model is a crucial step in successfully answering questions about the behavior of an experimental system. The model enables us to predict how the system will behave in response to certain stimuli, or helps us explain an observed behavior. Obtaining an adequate model of an experimental system is a laborious and error-prone task. Model construction involves the analysis of observations of system behavior, as well as the application of domain knowledge. In this process, different assumptions about the structure and behavior of the system can be made. Often, this results in a number of competing models, all justifiable by the available observations.

As an illustration of this point, consider a phytoplankton culture in a chemostat (Fig. 1). The chemostat is a type of bioreactor in which the nutrients required for cell growth are supplied continuously at a fixed rate to a culture vessel whose contents are continuously mixed. Medium, cells, and by-products are continually removed from the vessel, maintaining the culture in the chemostat chamber at a constant volume. Fig. 2 shows two rival models describing the nutrient-limited growth of phytoplankton in a chemostat. The Monod model [40] assumes that the consumed nutrient is transformed into biomass instantaneously. This assumption is expressed by the linear proportionality between the growth rate μ and the nutrient uptake rate ρ. The Droop model [12, 20] decouples specific growth rate from external nutrient concentration by introducing an intracellular store of nutrients. The specific growth rate μ is hypothesized to depend on a quantity q called the cell quota, the average amount of stored nutrients in each cell.

The problem of *model discrimination* arises when a set of alternative models of a system exists, each resting on different assumptions about its structure and behavior. In order to discriminate between the competing models, and identify which of them most adequately describes the actual situation, new observations have to be made. These can be obtained by performing additional experiments on the system. An experiment discriminates between the models, if the predictions of some of the models fit the newly-obtained data whereas the predictions of others do not. Since in real-life applications the number of experiments that can be performed may be quite large, and the cost of each of them considerable, it is important that the experiments be selected carefully. It is preferable to select experiments in such a way that the set of possible models is maximally reduced (*systematic* model discrimination) at minimal costs (*efficient* model discrimination).

The experiment that best discriminates between the models, the *optimal discriminatory experiment*, is not generally known. In fact, the only way to determine this experiment would be to perform all possible experiments, something we obviously want to avoid. Instead, we can predict the optimal discriminatory experiment by investigating the predictions of the system behavior derived from the competing models. It is intuitively clear that an experiment for which the model predictions are distant will have a higher chance of discriminating between the models than an experiment for which the predictions are close.

Model discrimination has received quite some attention in the statistical literature (see [23, 24] for reviews). The methods described deal with the problem of how to find an experiment that brings the system into an observable state for which the model predictions are as diverse as possible. This idea has been elaborated for the case of discrimination between mathematical models with an exactly-defined structure and precise numerical values for the system parameters.

In many situations, however, the formulation of a quantitative model is impeded by incomplete or imprecise information about the system structure and behavior. Incomplete or imprecise information may arise from inaccurate measurement instruments, inadequate understanding of the physical processes involved, inability to monitor the behavior of the system, *etc.* As a result, precise numerical values of the model parameters and exact functionals between the model variables cannot be determined. Biology and chemistry are examples of domains where this is often the case.

One way of dealing with incomplete and imprecise knowledge is to consider numerical intervals that bound parameter values, and to specify envelopes that bound unknown functions, giving rise to

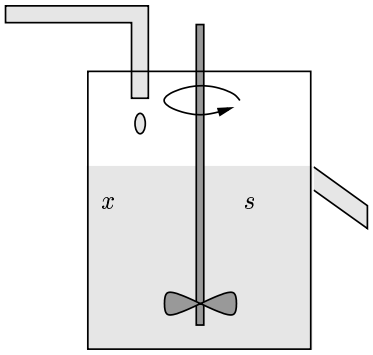


Figure 1: Schematic illustration of the chemostat.

$$\dot{x} = \mu(s)x - Dx$$
$$\dot{s} = D(s_{in} - s) - \rho(s)x$$
$$\mu(s) = \mu_{max}\frac{s}{s + k_s}$$
$$\rho(s) = \frac{1}{Y}\mu(s)$$

(Monod)

$$\dot{x} = \mu(q)x - Dx$$
$$\dot{q} = \rho(s) - \mu(q)q$$
$$\dot{s} = D(s_{in} - s) - \rho(s)x$$
$$\mu(q) = \bar{\mu}(1 - \frac{k_q}{q})$$
$$\rho(s) = \rho_{max}\frac{s}{s + k_s}$$

(Droop)

Figure 2: Monod and the Droop model describing the nutrient-limited growth of phytoplankton in the chemostat. x denotes the total amount of biomass per unit volume, s the concentration of remaining nutrient, q the internal cell quota. See Fig. 10 for the meaning and the units of the parameters in the models.

so-called *semi-quantitative models* [5, 31]. Semi-quantitative models are useful in many scientific and engineering domains, which often deal with parametric and functional tolerances described by intervals and bounding curves. The classical methods for model discrimination referred to above are unable to deal with such situations. This has motivated the work described in this paper.

We present a method for the systematic and efficient discrimination between semi-quantitative models through the selection of optimal discriminatory experiments [48, 49]. Experiment selection is based on an entropy criterion that generalizes upon previous work in statistics. The criterion estimates the discriminatory potential of the possible experiments by exploiting the predictions of the models obtained through semi-quantitative simulation, and formalizes the intuition that experiments for which the model predictions are more distant have a higher potential to discriminate between the alternatives. The algorithm for model discrimination is formalized as an iterative process, starting with a set of competing models to which initial probabilities have been assigned. At every step, the experiment with the highest discriminatory potential is determined, this experiment is executed, and the experimental outcome is then used to recompute the probabilities of the competing models. This process continues until one of the models has reached a cut-off probability, all models have zero probabilities, or the possible experiments have been exhausted.

The ability of the method to efficiently discriminate between competing models is demonstrated on a real application involving a set of complex models of an incompletely-specified experimental system, costly experiments, and imprecise data. The example concerns a problem in population biology, a domain where model discrimination is a key issue since most of the models are derived from empirical considerations and have rarely been appropriately validated, thus giving rise to several competing models. We will be especially concerned with the selection of experiments to discriminate between

competing models of nutrient-limited phytoplankton growth in a chemostat. At least four models have been used to describe algal growth and substrate consumption and it is not clear which one is the best. The choice of optimal discriminatory experiments is critical in this application, since the doubling time of the population is of the order of two days, which means that the experiments take several weeks to complete. Semi-quantitative models are appropriate here because, as is the case for most biological systems, the available data is noisy and fragmentary.

The article is organized as follows. In the next section, some basic concepts are defined, and the method for model discrimination is informally introduced. The method is based on a generalized entropy criterion (Sec. 3) that guides the selection of optimal discriminatory experiments. In Sec. 4, the algorithm for model discrimination is presented in detail. The results of the application of the method to the modeling of the nutrient-limited phytoplankton growth in a chemostat are presented in Sec. 5. In Sec. 6, the method for model discrimination is discussed in the context of related work, while the last section summarizes what has been achieved and indicates directions for further work.

2 Basic concepts and outline

This section provides the conceptual basis of the method for model discrimination. In particular, the concepts of experimental system, experiment, and competing model are introduced and explained. Also, an outline of the method is given.

2.1 Experimental systems and experiments

Experimental systems are (physical, chemical, biological) systems that are created and sustained in an experimental *set-up* [16, 17]. Consider, for instance, phytoplankton growth in a chemostat. The experimental set-up consists of a phytoplankton culture in a vessel with a continuously-supplied growth medium, where various factors, such as the light intensity, the temperature, and the concentration of the feeding substrates can be controlled. The experimental set-up thus defines the possible *experimental conditions* under which the system can be investigated (*i.e.*, the ranges of temperatures, light intensities, substrate concentrations, *etc.*).

Characteristic properties of experimental systems are that their behavior can be *controlled* and *observed*. Control is achieved by maintaining the structure of the system and imposing the experimental conditions. Observations allow properties of the system to be determined. An *experiment* is an action, in which specific experimental conditions are chosen, and the behavior of the system under these conditions is observed. In the context of phytoplankton growth, an experiment might for example consist of observing the evolution of the phytoplankton biovolume from given initial conditions, while the concentration of the feeding substrate is fixed at 100 μmol/l.

2.2 Competing models of an experimental system

Generally speaking, a *model* of an experimental system is a symbolic structure consisting of a set of (state and input) variables, parameters, and their functional relations. The variables and parameters represent quantities of the system, while the functional relations describe interactions between the system components as well as interactions of the system with its environment. A model is usually valid within only a certain range of experimental conditions that can be realized in the experimental set-up. For instance, a model may be valid for phytoplankton growth at high light intensities only. In what follows we assume that the domain of validity of the models includes in the range of experimental conditions relevant to the problem being studied.

The *adequacy* of a model refers to the extent to which its predictions are in agreement with the behavior of the system in the range of experimental conditions of interest. A model is said to be *correct*, if it is maximally adequate, that is, if it is consistent with the behavior of the system in the

whole range of experimental conditions. This definition of the correctness suggests which criteria a model has to satisfy so as to qualify as a *candidate model* of the experimental system. In order to be able to compare the predictions of the model with the behavior of the system, we require that (some of) the model variables correspond to observed, that is, measured system quantities. Furthermore, we demand that the predictions obtained from the model are consistent with all available observations.

Def. 1 (Candidate model) A model m with variables $\boldsymbol{q}$ is a candidate model of a given experimental system, if the following conditions hold:

1. $\boldsymbol{q}_{obs} \cap \boldsymbol{q} \neq \emptyset$, where the vector $\boldsymbol{q}_{obs}$ denotes the observed quantities of the experimental system;
2. the predicted values of $\boldsymbol{q}_{obs}$ are consistent with all available observations.

Consider again the example of phytoplankton growth in a chemostat. Among the quantities measured during the experiments are the amount of biomass and the concentration of the remaining limiting nutrient. The Monod model contains variables corresponding to these quantities (called x and s, respectively). If the predicted values for x and s are consistent with the observations, then according to Def. 1, the Monod model qualifies as a candidate model of the experimental system.

The problem of model discrimination arises when there are several candidate models of an experimental system, making different assumptions on the structure and behavior of the experimental system. For the models to be *competing*, they must have common observed variables. In addition, the candidate models have to yield different predictions for these variables in the range of experimental conditions in which we study the system. These criteria are summarized in the following definition.

Def. 2 (Competing models) Let $m_1, \ldots, m_l$ be candidate models with variables $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_l$, respectively. Furthermore, let $\boldsymbol{q}_{int} = \boldsymbol{q}_{obs} \cap \boldsymbol{q}_1 \cap \ldots \cap \boldsymbol{q}_l$. The models are competing, if

1. $\boldsymbol{q}_{int} \neq \emptyset$;
2. for every pair of models m_i and m_j, $1 \leq i, j \leq l$, $i \neq j$, there are some experimental conditions for which the values of $\boldsymbol{q}_{int}$ are inconsistent.

As can be seen from Fig. 2, the Monod and the Droop models share two variables corresponding to observed quantities, namely the amount of biomass x and the concentration of the limiting nutrient s. Furthermore, the simulation results of Fig. 5 show that for some experimental conditions, the models predict a different time evolution of these variables. Hence, according to Def. 2, the two models are competing.

Let $M = \{m_1, \ldots, m_l\}$ be a set of competing models. For each $m_i \in M$, we define $p(m_i)$ to stand for the probability that m_i is a correct model of the experimental system in the range of experimental conditions of interest. Notice that it follows from the second criterion in Def. 2 that at most one model in M can be correct (*i.e.*, maximally adequate). *A priori* estimates of the probabilities can be determined from available experimental data. If no such data exist, we assume that the models are equiprobable *a priori*. When new data become available, the probability estimates can be updated according to the match between the model predictions and the observations.

2.3 Experiment selection for model discrimination

Discrimination between the competing models in M can be achieved by selecting and performing experiments from a predetermined set of possible experiments E. The experiments in E are assumed to have been chosen from within the range of experimental conditions relevant to the problem under study. Recall from Sec. 2.2 that the competing models in M are valid under these experimental conditions.

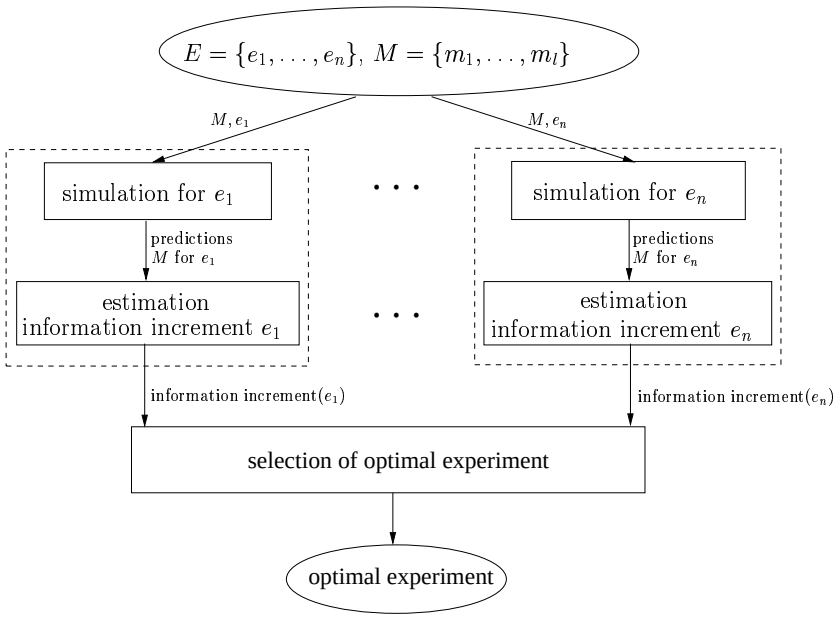


Figure 3: Schematic overview of the selection of experiments for model discrimination.

An important question is what experiments would be relevant in an experimental set-up, as the number of possibilities might be infinite. In practice, the set of possible experiments will be restricted by various constraints. First, not everything is feasible in a given experimental set-up. Too high a pressure in a reactor may cause an explosion, while too low a temperature may not be realizable with the available equipment. Another set of constraints is imposed by the requirement that the experiments be chosen within the range of experimental conditions of interest. For instance, if we study phytoplankton growth under high light intensities only, and our models are valid under these conditions only, it does not make much sense to perform discriminatory experiments at low light intensities. In this study, the set of possible experiments will be assumed given and in accordance with the above constraints.

For each $e_j \in E$, every $m_i \in M$ has to be simulated with the appropriate parameter values and initial conditions, in order to determine the predictions of m_i for the behavior of the system in e_j. The predictions of the competing models, together with the current probability estimates of the models, are then used to compute the expected information increment of e_j (Fig. 3). The expected information increment is a criterion that measures the discriminatory potential of an experiment. Intuitively, an experiment for which the model predictions are more distant can be expected to lead to better discrimination, as the experimental outcome will generally agree with fewer of the predictions. If all models give rise to the same predictions for a certain e_j, then the experiment will not be able to discriminate between the models. In Sec. 3 this intuition will be elaborated by means of an approach based on concepts from information theory.

Once the optimal discriminatory experiment has been determined, it is executed. The observations thus obtained are used to adjust the probability estimates $p(m_i)$ of the models $m_i \in M$. The experiment discriminates between the models in M, if the new probability estimates of the models are different. Optimal discrimination will be achieved if the observations are consistent with the predictions of only a single model. If the observations of the behavior of the system are consistent with the predictions of several models, further experiments need to be selected and carried out. In this way, the problem of model discrimination becomes an iterative process. The procedure of selecting an experiment, performing this experiment, and updating the probability estimates of the competing models (Fig. 4) is repeated until one of the following happens: a model has a sufficiently high probability; all models

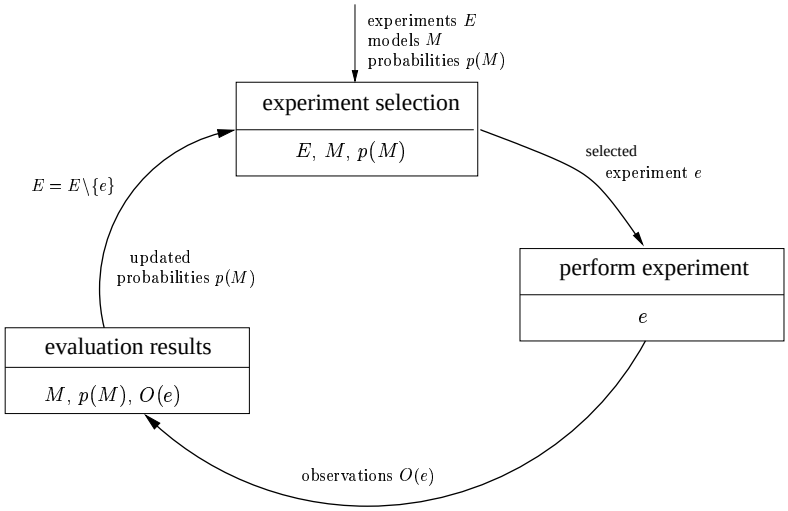


Figure 4: Model discrimination as an iterative process.

in M have zero probability (because they are inconsistent with the observations); or the set of possible experiments E is exhausted.

3 Semi-quantitative modeling and simulation

A key step in model discrimination is the derivation of predictions from competing models (Fig. 3). This is achieved through simulation. As information about experimental systems is often incomplete and imprecise, traditional numerical simulation techniques are difficult to apply. Instead, we will use semi-quantitative simulation techniques that require weaker quantitative information, in particular numerical bounds on parameter values and functions.

A variety of *semi-quantitative* or *interval simulation techniques* have been developed (reviewed in [48]). In this paper, we will employ techniques that have been developed in the field of qualitative reasoning, in particular the technique QSIM [34], and its extensions Q2 [34, 35] and Q3 [5]. Essentially, semi-quantitative simulation consists of refining the predictions of the qualitative simulator QSIM, through the integration of quantitative information by means of Q2 and Q3. Although convenient for our purpose, it should be noted that the model discrimination method presented in this paper does not in any way depend on the particular set of simulation techniques chosen. The only requirement posed by our method is that the simulation techniques yield interval predictions for the temporal evolution of the variables.

In QSIM, the domain of a variable x is defined by its *quantity space*. The quantity space is a set of totally ordered *landmark values* $l_1 < \ldots < l_k$ denoting qualitatively-significant values of the variable. The *qualitative value* $QV(x, t_j)$ of x at a time-point t_j, or the qualitative value $QV(x, t_j, t_{j+1})$ between two time-points t_j and t_{j+1}, is expressed in terms of the landmarks in its quantity space and its direction of change (*std*, *inc*, or *dec*). The *qualitative state* of a system at a time-point t_j, or between two time-points t_j and t_{j+1}, consists of a tuple of qualitative values, one for each variable of the system, while the *qualitative behavior* of the system on a time-interval $[a, b]$ is given by the sequence of qualitative states traversed by the system on that interval (see [34] for details).

QSIM models experimental systems by means of *qualitative differential equations (QDEs)*, which are qualitative abstractions of ordinary differential equations (ODEs). More precisely, a QDE consists of a set of constraints on the qualitative values of the system variables. A QDE is abstracted from an ODE by first decomposing the latter into a set of basic mathematical equations, and then mapping the

equations to corresponding constraints between the qualitative values of the variables [34]. Consider, for instance, the Monod model given in Fig. 2. The first equation $\dot{x} = (\mu(s) - D)\, x$ can be decomposed into the following basic equations:

$$\dot{x} = a,\ a = b\, x,\ b = \mu - D,\ \dot{D} = 0,$$

where a and b represent auxiliary variables. Next, these equations can be mapped to constraints between the qualitative values of the variables:

$$\mathrm{D/DT}(a, x), \mathrm{MULT}(b, x, a), \mathrm{ADD}(b, D, \mu), \mathrm{CONSTANT}(D).$$

The above concepts are extended in the framework of semi-quantitative simulation by adding numerical information. In particular, the landmarks $l_1, \ldots, l_k$ of a variable x are augmented with numerical ranges. This gives rise to *semi-quantitative values*, *states*, and *behaviors* [48]. By enhancing QDEs with quantitative information, in the form of numerical ranges for parameters and initial conditions, as well as numerical envelopes for functions, *semi-quantitative differential equations (SQDEs)* are obtained. For instance, the models in Fig. 2 can be converted to SQDEs by specifying the following intervals for the parameters and the initial conditions:

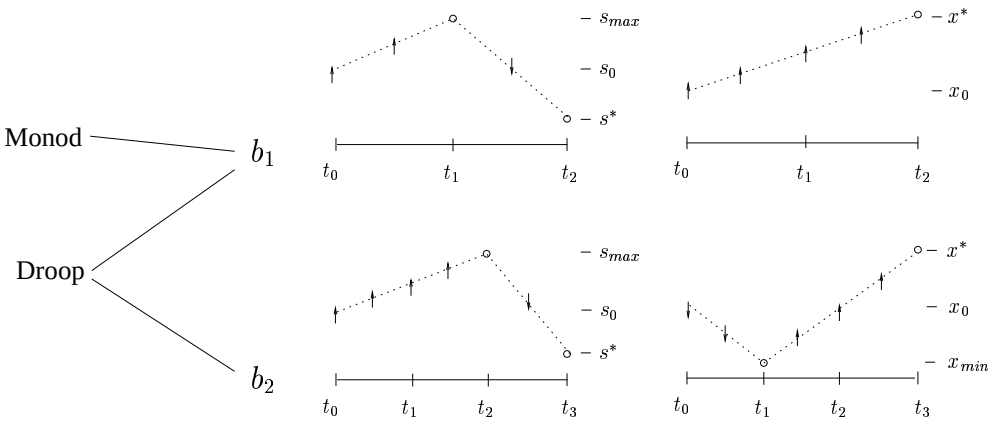
$$\begin{array}{l} \mu_{max} \in [1.2, 1.6],\ k_s \in [0.01, 0.2],\ Y \in [0.15, 0.6],\ s_{in} \in [80, 120], \\ \bar{\mu} \in [1.7, 2.3],\ \rho_{max} \in [9.25, 9.55],\ k_q \in [1.6, 2.0],\ D \in [0.38, 0.42], \\ x_0 \in [0.088, 0.165],\ s_0 \in [44, 46],\ q_0 \in [1.6, 7.62]. \end{array} \quad (1)$$

The QSIM algorithm predicts all possible qualitative behaviors of an experimental system based on its QDE description. Q2 and Q3 exploit the semi-quantitative information in an SQDE to refine the qualitative behavior tree produced by QSIM. More specifically, they rule out qualitative behaviors, or transform qualitative behaviors into semi-quantitative behaviors, by calculating numerical ranges for the landmarks of the variables. The behaviors resulting from the simulation of the Monod and Droop models (Fig. 2) with numerical ranges (1) are shown in Fig. 5.

QSIM, Q2, and Q3 have been proven *sound* [5, 34, 35]. That is, all possible semi-quanti-tative behaviors consistent with an SQDE are generated by the simulation algorithms. On the other hand, QSIM and its extensions are *incomplete*, in that spurious qualitative behaviors may be generated or landmark ranges overestimated. Given a predicted semi-quantitative behavior, the landmark ranges can be refined after execution of the experiment by integrating measurements of observed variables [21, 32]. Measurements obtained in previous experiments can be integrated as well, by deriving additional constraints from a comparison of the models describing the system in different experiments. This is achieved by means of SQCA, a technique for the *semi-quantitative comparative analysis* of dynamical systems described by SQDEs [48, 50].

4 Criterion for ranking experiments

Different criteria for evaluating the discriminatory potential of an experiment have been proposed in the statistical literature (*e.g.*, [1, 22, 23, 24, 27]). However, the methods previously proposed are limited to the case of numerical models with a precisely-defined structure that give rise to point predictions. In this section we propose a criterion for the selection of experiments that guides the discrimination of semi-quantitative competing models [48, 49]. The criterion is formalized by using concepts from information theory and generalizes upon previous work in statistics.



	s_{max}	$s^*(\times 10^{-3})$	x_{min}	x^*
Monod b_1	$[47.75, 118.36]$	$[3.11, 107.69]$	–	$[11.98, 72.0]$
Droop b_1	$[44.0, 115.17]$	$[0.82, 27.43]$	–	$[30.11, 62.61]$
Droop b_2	$[44.0, 118.09]$	$[0.82, 27.43]$	$[0.076, 0.165]$	$[30.11, 62.61]$

	t_1	t_2	t_3
Monod b_1	$[0.71, 147.09]$	$[1.21, \infty]$	–
Droop b_1	$[0.53, \infty]$	$[1.16, \infty]$	–
Droop b_2	$[0, 0.15]$	$[0.52, \infty]$	$[1.17, \infty]$

Figure 5: Semi-quantitative behaviors predicted by the Monod and Droop models with the numerical ranges (1). The Monod model predicts a single semi-quantitative behavior in which the biomass x increases asymptotically to its equilibrium x^*, and s passes through a maximum before reaching its equilibrium s^*. The Droop model predicts two semi-quantitative behaviors. The first behavior is qualitatively equivalent to the behavior predicted by the Monod model. In the second behavior, the biomass x first reaches a minimum, followed by a maximum of s, and then the equilibrium is approached. The table gives the intervals for the landmark values in the graphs.

4.1 Expected information increment

Let M be a set of competing models of an experimental system with *a priori* probabilities $p(m_i)$ for every $m_i \in M$. We make the common assumption that M is complete:

$$\sum_{m_i \in M} p(m_i) = 1. \tag{2}$$

That is, we assume that M contains a correct model, in the sense of Sec. 2.2. Because the models are competing, M contains exactly one correct model. Although (2) is a strong assumption, the practical consequences are limited, because its violation can be tested.

Consider the set of possible experiments E, and let $e \in E$ be an experiment in which behavior b^e is observed. b^e is characterized by a set of *behavioral features*, that is, landmarks representing, for instance, minima, maxima, and equilibria of the system quantities. For clarity of presentation, we will assume for the moment that b^e is characterized by a single feature. The measurement of the behavioral feature in e is assumed to be a random variable Y^e with mean y^e and a probability density function $h(y|y^e, b^e)$ of known analytic form (uniform, normal, ...) [23]. That is,

$$E(Y^e) = \int_{y \in D(b^e)} y\, h(y|y^e, b^e)\, dy = y^e,$$

where $D(b^e) \subseteq \mathbb{R}$ denotes the domain of the behavioral feature of b^e.

A standard measure in information theory is the *information increment* of an experiment [11, 23]. The information increment of e is defined as the difference in entropy before and after the execution of the experiment:

$$\Delta H(y^e, b^e) = -\sum_{m_i \in M} p(m_i) \ln p(m_i) + \sum_{m_i \in M} p^e(m_i \mid y^e, b^e) \ln p^e(m_i \mid y^e, b^e), \tag{3}$$

with $p^e(m_i \mid y^e, b^e)$ the *a posteriori* probability of m_i.

ΔH reaches its maximum when all posterior probabilities but one are zero. That is, when the observations obtained in e confirm the predictions of a single model. On the other hand, a minimal value is attained, when all posterior probabilities are equal.

Since the *a posteriori* probabilities of the models depend on the outcome of an experiment that has not yet been performed, ΔH cannot be computed directly. Instead, we can compute its expected value, or the *expected information increment* ΔJ of e. By definition,

$$\Delta J(e) = E[\Delta H(y^e, b^e)] = \sum_{b \in B(e)} \int_{y \in D(b)} \Delta H(y, b)\, f^e(y, b)\, dy, \tag{4}$$

where $B(e)$ is the set of possible qualitative behaviors b of the system, as predicted by the competing models, $D(b)$ the domain of the behavioral feature of b, and $f^e(y, b)$ the joint probability density function (pdf) of y and b. The function $f^e(y, b)$ is defined as follows:

$$f^e(y, b) = f^e(y|b)\, p^e(b), \tag{5}$$

where $f^e(y|b)$ is the conditional pdf of y, given that b is the behavior of the system in experiment e, and $p^e(b)$ the probability that the system exhibits behavior b in e. According to the law of total probability, we can express the probability of b as a weighted sum of the conditional behavior probabilities $p^e(b|m_i)$:

$$p^e(b) = \sum_{m_i \in M} p^e(b|m_i)\, p(m_i), \tag{6}$$

where $p^e(b|m_i)$ is the probability that the system behavior observed in e is b, provided m_i is the correct model of the system. Similarly, according to the same law, we can express the conditional pdf $f^e(y|b)$ as follows:

$$f^e(y|b) = \int_{x \in D(b)} h^e(y|x, b)\, g^e(x|b)\, dx, \tag{7}$$

where $h^e(y|x, b)$ is the conditional pdf of the measurement of the behavioral feature in e, and $g^e(x|b)$ the pdf of the mean x. An expression for $g^e(x|b)$ can be obtained from the predictions of the competing models. More specifically, the law of total probability gives us:

$$g^e(x|b) = \sum_{m_i \in M} g^e(x|b, m_i)\, p^e(m_i|b), \tag{8}$$

where $g^e(x|b, m_i)$ is the model-specific pdf of the mean given b is the system behavior in e, and m_i the correct model of the system.

Now, substituting (8) into (7) gives

$$f^e(y|b) = \sum_{m_i \in M} \int_{x \in D(b)} h^e(y|x, b)\, g^e(x|b, m_i)\, p^e(m_i|b)\, dx. \tag{9}$$

From (9), for a fixed m_i, we obtain

$$f^e(y|b, m_i) = \int_{x \in D(b)} h^e(y|x, b)\, g^e(x|b, m_i)\, dx. \tag{10}$$

The model-specific pdfs $h^e(y|x,b)$ and $g^e(x|b,m_i)$, and the conditional behavior probabilities $p^e(b|m_i)$, satisfy the following normalization conditions:

$$\int_{y\in D(b)} h^e(y|x,b)\,dy = 1, \tag{11}$$

$$\int_{x\in D(b)} g^e(x|b,m_i)\,dx = 1, \tag{12}$$

$$\sum_{b\in B_i(e)} p^e(b|m_i) = 1, \tag{13}$$

where $B_i(e)$ is the set of qualitative behaviors of the system in e, as predicted by m_i. With (11) and (12) it follows that

$$\int_{y\in D(b)} f^e(y|b,m_i)\,dy = 1. \tag{14}$$

Sec. 4.2 and Sec. 4.3 consider the determination of the model-specific pdfs and the conditional behavior probabilities in detail.

In the following proposition, the probability density function $f^e(y|b)$, the probability $p^e(b)$, and the normalization conditions (2), (12), (13), and (14) are used to derive an expression for the expected information increment ΔJ.

Prop. 1 The expected information increment of e is given by

$$\Delta J(e) = \sum_{m_i\in M} p(m_i) \sum_{b\in B_i(e)} p^e(b|m_i) \int_{y\in D(b)} f^e(y|b,m_i) \ln \frac{f^e(y|b,m_i)p^e(b|m_i)}{f^e(y|b)p^e(b)} dy. \tag{15}$$

Proof. Using Bayes' rule we obtain for the *a posteriori* probability $p^e(m_i|b)$:

$$p^e(m_i|b) = \frac{p^e(b|m_i)\,p(m_i)}{p^e(b)}.$$

Substituting $p^e(m_i|b)$ in (8), the following expression for the probability density function $g^e(x|b)$ is obtained:

$$g^e(x|b) = \frac{\sum_{m_i\in M} g^e(x|b,m_i)\,p^e(b|m_i)\,p(m_i)}{p^e(b)}.$$

Substitution of this expression into (7), together with (10), yields

$$\begin{aligned} f^e(y|b) &= \int_{x\in D(b)} h^e(y|x,b)\,\frac{\sum_{m_i\in M} g^e(x|b,m_i)\,p^e(b|m_i)\,p(m_i)}{p^e(b)}\,dx \\ &= \frac{\sum_{m_i\in M} f^e(y|b,m_i)\,p^e(b|m_i)\,p(m_i)}{p^e(b)}. \end{aligned} \tag{16}$$

By substituting the definition of ΔH into (4), and applying (5), we get

$$\begin{aligned} \Delta J(e) = \sum_{b\in B(e)} \int_{y\in D(b)} & \\ \left\{ \sum_{m_i\in M} p^e(m_i|y,b) \ln p^e(m_i|y,b) - \sum_{m_i\in M} p(m_i) \ln p(m_i) \right\} & f^e(y|b)\,p^e(b)\,dy, \end{aligned} \tag{17}$$

where

$$p^e(m_i|y,b) = \frac{f^e(y,b|m_i)\,p(m_i)}{f^e(y,b)} = \frac{f^e(y|b,m_i)\,p^e(b|m_i)\,p(m_i)}{f^e(y|b)\,p^e(b)}, \tag{18}$$

via Bayes' rule and (5). Combining (17) and (18), we obtain:

$$\begin{aligned}
\Delta J(e) = & \sum_{b\in B(e)} \int_{y\in D(b)} \sum_{m_i\in M} \\
& p(m_i)\left\{\frac{f^e(y|b,m_i)\,p^e(b|m_i)}{f^e(y|b)\,p^e(b)} \ln \frac{f^e(y|b,m_i)\,p^e(b|m_i)\,p(m_i)}{f^e(y|b)\,p^e(b)} - \ln p(m_i)\right\} f^e(y|b)\,p^e(b)\,dy \\
= & \sum_{m_i\in M} p(m_i) \sum_{b\in B_i(e)} p^e(b|m_i) \int_{y\in D(b)} f^e(y|b,m_i) \ln \frac{f^e(y|b,m_i)\,p^e(b|m_i)}{f^e(y|b)\,p^e(b)}\,dy \\
& + \sum_{m_i\in M} p(m_i)\ln p(m_i) \sum_{b\in B_i(e)} p^e(b|m_i) \int_{y\in D(b)} f^e(y|b,m_i)\,dy \\
& - \sum_{m_i\in M} p(m_i)\ln p(m_i) \sum_{b\in B_i(e)} \int_{y\in D(b)} f^e(y|b)\,p^e(b)\,dy.
\end{aligned}$$

Using the normalizations (2), (12), (13), and (14), we obtain

$$\sum_{m_i\in M} p(m_i)\ln p(m_i) \sum_{b\in B_i(e)} p^e(b|m_i) \int_{y\in D(b)} f^e(y|b,m_i)\,dy = \sum_{m_i\in M} p(m_i)\ln p(m_i),$$

$$\sum_{m_i\in M} p(m_i)\ln p(m_i) \sum_{b\in B_i(e)} \int_{y\in D(b)} f^e(y|b)\,p^e(b)dy = \sum_{m_i\in M} p(m_i)\ln p(m_i).$$

With these equations, the second part of the expression for ΔJ becomes 0, and (15) is obtained.

□

The criterion ΔJ ranks the experiments in E according to their expected informativeness. The optimal discriminatory experiment will be the experiment that is *expected* to be most informative, or the *most informative experiment* for short. That is, the experiment for which $\Delta J(e)$ is maximal.

Occasionally, the model predictions can lead to a simplified expression for ΔJ. The following corollary of Prop. 1 provides an expression for ΔJ when all models predict the same qualitative behavior [51].

Corr. 1 Let all models in M predict the same qualitative behavior b of the system in e. Then (15) takes the form

$$\Delta J(e) = \sum_{m_i\in M} p(m_i) \int_{y\in D(b)} f^e(y|b,m_i) \ln \frac{f^e(y|b,m_i)}{f^e(y|b)}\,dy. \tag{19}$$

Proof. If all models predict the same qualitative behavior, (15) can be rewritten as

$$\Delta J(e) = \sum_{m_i\in M} p(m_i)\,p^e(b|m_i) \int_{y\in D(b)} f^e(y|b,m_i) \ln \frac{f^e(y|b,m_i)\,p^e(b|m_i)}{f^e(y|b)\,p^e(b)}\,dy.$$

Taking into account that $p^e(b|m_i) = 1$ for all $m_i \in M$, and

$$p^e(b) = \sum_{m_i\in M} p^e(b|m_i)\,p(m_i) = \sum_{m_i\in M} p(m_i) = 1,$$

the result (19) is obtained.

□

The following corollary derives an expression for ΔJ for the reverse situation: when each model predicts a different set of qualitative behaviors.

Corr. 2 Assume that for a given e, each model $m_i \in M$ predicts a different set of qualitative behaviors. Then,

$$\Delta J(e) = -\sum_{m_i \in M} p(m_i) \ln p(m_i). \tag{20}$$

Proof. From the given assumptions we have $p^e(b|m_k) = 0$, if $b \in B_i(e)$ and $k \neq i$. This fact, together with the expression for $f^e(y|b)$ given in (16), yields

$$f^e(y|b)\, p^e(b) = \sum_{m_k \in M} f^e(y|b, m_k)\, p^e(b|m_k)\, p(m_k) = f^e(y|b, m_i)\, p^e(b|m_i)\, p(m_i),$$

for all $b \in B_i(e)$. Taking this result into account, (15) takes the form

$$\Delta J(e) = \sum_{m_i \in M} p(m_i) \sum_{b \in B_i(e)} p^e(b|m_i) \int_{y \in D(b)} f^e(y|b, m_i) \ln \frac{1}{p(m_i)}\, dy,$$

or equivalently,

$$\Delta J(e) = -\sum_{m_i \in M} p(m_i) \ln p(m_i) \sum_{b \in B_i(e)} p^e(b|m_i) \int_{y \in D(b)} f^e(y|b, m_i)\, dy.$$

By applying the normalization conditions (13) and (14), expression (20) is obtained.

□

Note that (20) indicates the maximum value that $\Delta J(e)$ can take. This conforms to the intuition that e will be the optimal discriminatory experiment since its outcome is guaranteed to be consistent with the predictions of at most one model.

The criterion (15) is easily generalizable to the case that each behavior b is characterized by more than one feature. In this case we have to substitute the probability distributions by joint probability distributions, and the integral by a multiple integral for all behavioral features.

4.2 Estimation of behavior probabilities

In order to compute $\Delta J(e)$, the conditional behavior probabilities $p^e(b|m_i)$ must be known. Estimates of these probabilities have been obtained by means of the following approach. Let θ_l, with $\theta_l \in \Theta_l$, be a parameter or an initial condition of model m_i, and let $\Theta_{lj} \subseteq \Theta_l$ be the interval value of θ_l for which behavior $b_j \in B_i(e)$ is obtained. The intervals Θ_{lj} are obtained directly by simulation with Q2 and Q3. Assume the probability distributions of the parameters θ_l are given. For each $b_j \in B_i(e)$ we define:

$$r^e(b_j|m_i) = \prod_{l \in L} p(\theta_l \in \Theta_{lj}),$$

where $p(\theta_l \in \Theta_{lj})$ is the probability that the value of θ_l lies in the interval Θ_{lj}, and L the set of indices of parameters and initial conditions.

The conditional probability of behavior b_j is now estimated by normalizing the $r^e(b_j|m_i)$s:

$$p^e(b_j|m_i) = \frac{r^e(b_j|m_i)}{\sum_{b_k \in B_i(e)} r^e(b_k|m_i)}. \tag{21}$$

If the probability distributions of the θ_ls are unknown, we assume that the θ_ls are uniformly distributed. That is,

$$p(\theta_l \in \Theta_{lj}) = \frac{\mid \Theta_{lj} \mid}{\mid \Theta_l \mid},$$

where $\mid \cdot \mid$ denotes interval length. In this case, $p^e(b_j|m_i)$ is given by the fraction of the parameter volume covered by b_j.

Consider the behaviors b_1 and b_2 in Fig. 5, generated by simulation of the Droop model. b_1 is obtained for $q_0 \in [1.92, 7.62]$, while b_2 is obtained for $q_0 \in [1.6, 2.66]$. If we assume that q_0 is uniformly distributed in the interval $[1.6, 7.62]$, then the procedure outlined above gives us $r(b_1|\mathrm{D}) = 0.9468$, $r(b_2|\mathrm{D}) = 0.1761$, where D stands for the Droop model. After normalization, we obtain the probability estimates $p(b_1|\mathrm{D}) = 0.8432$, $p(b_2|\mathrm{D}) = 0.1568$.

Alternative approaches to estimating the probabilities of qualitative behaviors have been proposed by Berleant *et al.* [4] and Leitch and Shen [37]. In [4] the interval values of the parameters and the initial conditions are used to derive a lower and an upper bound of the probability of a qualitative behavior. We have not adopted this approach here, because the computation of the value of the expected information increment requires that the conditional behavior probabilities are real values. Leitch and Shen [37] have proposed an algorithm to prioritize the qualitative behaviors of a model. The method uses imprecise numerical information in the form of fuzzy numbers to define distance between successor states in the qualitative behavior tree. Each state is given a priority label according to the value of this distance. The priority of a behavior is estimated on the basis of the priorities of its states. For our purposes, a disadvantage of this algorithm is that it only orders the behaviors (*i.e.*, b_1 is more likely to occur than b_2), but does not derive a quantitative estimate of their relative priorities (*i.e.*, by how much is b_1 more likely to occur than b_2).

4.3 Estimation of probability density functions

The computation of $\Delta J(e)$ also requires the density functions $f^e(y|b, m_i)$ of the measurements of the behavioral features, where

$$f^e(y|b, m_i) = \int_{x \in D(b)} h^e(y|x, b)\, g^e(x|b, m_i)\, dx. \tag{22}$$

The conditional pdfs $h^e(y|x, b)$ are assumed given. For the moment, we will assume that the behavioral features are uniformly distributed on the confidence interval of the measurement. That is, $h^e(y|x, b) = 1/\varepsilon$, for $y \in [x - \varepsilon/2, x + \varepsilon/2]$, with ε the size of the confidence interval.

In order to determine $f^e(y|b, m_i)$, and hence $\Delta J(e)$, the model-specific pdfs $g^e(x|b, m_i)$ of the means of the behavioral features have to be determined. The predictions of the competing models for the values of the behavioral features, obtained by means of QSIM, Q2, and Q3, have the form of intervals (Sec. 3). Since nothing is known about the distribution within a predicted interval, we assume a uniform distribution. More specifically, if $V_i^e \subseteq D(b)$ is the interval predicted by m_i for the behavioral feature of b, then $g^e(x|b, m_i) = 1/|V_i^e|$ for $x \in V_i^e$, and 0 otherwise.

Given these assumptions, the integral in the definition of $f^e(y|b, m_i)$ can be computed exactly. More specifically, we have

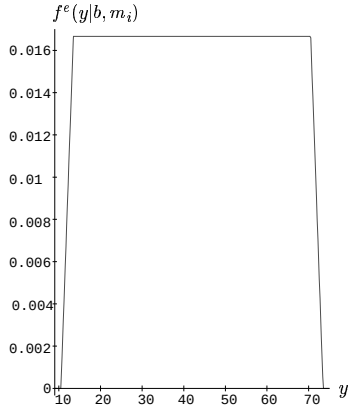


Figure 6: Plot of the function $f^e(y|b,m_i)$ for the behavioral feature x^* (equilibrium amount of biomass) of the behavior b_1 predicted by the Monod model (Fig. 5). The predicted interval for the equilibrium x^* is $[11.98, 71.99]$. (The size of the confidence interval ε has been set equal to 3.)

$$f^e(y|b,m_i) = \begin{cases} \dfrac{y - \underline{V_i^e} + \varepsilon/2}{\varepsilon |V_i^e|} & , y \in [\underline{V_i^e} - \varepsilon/2, \underline{V_i^e} + \varepsilon/2], \\ \dfrac{1}{|V_i^e|} & , y \in [\underline{V_i^e} + \varepsilon/2, \overline{V_i^e} - \varepsilon/2], \\ \dfrac{-y + \overline{V_i^e} + \varepsilon/2}{\varepsilon |V_i^e|} & , y \in [\overline{V_i^e} - \varepsilon/2, \overline{V_i^e} + \varepsilon/2], \\ 0 & , y \notin [\underline{V_i^e} - \varepsilon/2, \overline{V_i^e} + \varepsilon/2], \end{cases} \tag{23}$$

where $\underline{V_i^e}$ and $\overline{V_i^e}$ are the lower and the upper bound of the interval prediction V_i^e, respectively. Fig. 6 shows the estimate of the model-specific pdf of the behavioral feature x^*, the equilibrium value of the amount of biomass in behavior b_1 as predicted by the Monod model (Fig. 5).

The expression for $f^e(y|b,m_i)$ defined above assumes that the mean of the value of the behavioral feature is uniformly distributed in the predicted interval. Better estimations of the probability density functions $f^e(y|b,m_i)$ can be obtained in the following way. Let θ be a parameter or an initial condition of m_i, and $\theta \in \Theta$. We divide Θ into subintervals $\Theta_j \subseteq \Theta$ of equal length, for each of which we specify the probability $p(\theta \in \Theta_j)$. This results in an approximation of the probability distribution of θ in Θ. Obviously, if we do not know anything about the probability distributions of the parameters and initial conditions, then we set $p(\theta \in \Theta_j) = p(\theta \in \Theta_k)$ for all j, k. We now simulate m_i for each of the subintervals Θ_j. As a result, we obtain a set of (usually overlapping) intervals $V_{ij}^e \subseteq V_i^e$ for the value of the behavioral feature. For each of the predicted intervals V_{ij}^e we again assume uniform distribution of the mean. That is, $g_j^e(x|b,m_i) = 1/|V_{ij}^e|$ for $x \in V_{ij}^e$, and 0 otherwise. This results in the following estimate of the model-specific pdf $g^e(x|b,m_i)$:

$$g^e(x|b,m_i) = \sum_{\Theta_j} p(\theta \in \Theta_j) g_j^e(x|b,m_i).$$

By subdividing Θ into ever finer subintervals Θ_j, we obtain progressively better estimates of $g^e(x|b,m_i)$, and hence of $f^e(y|b,m_i)$.

Consider, for instance, the Monod model given in Fig. 2. Assume the parameters Y, s_{in}, m, and k_s are distributed in the corresponding intervals (1) as shown in Fig. 7(a). Every interval has been subdivided into a number of subintervals and the Monod model has been simulated for each combination of the subintervals. The resulting predictions have been used to estimate the model-specific pdf $g^e(x|b,m_i)$ of the equilibrium value x^* of the biomass. In comparison with Fig. 6, where

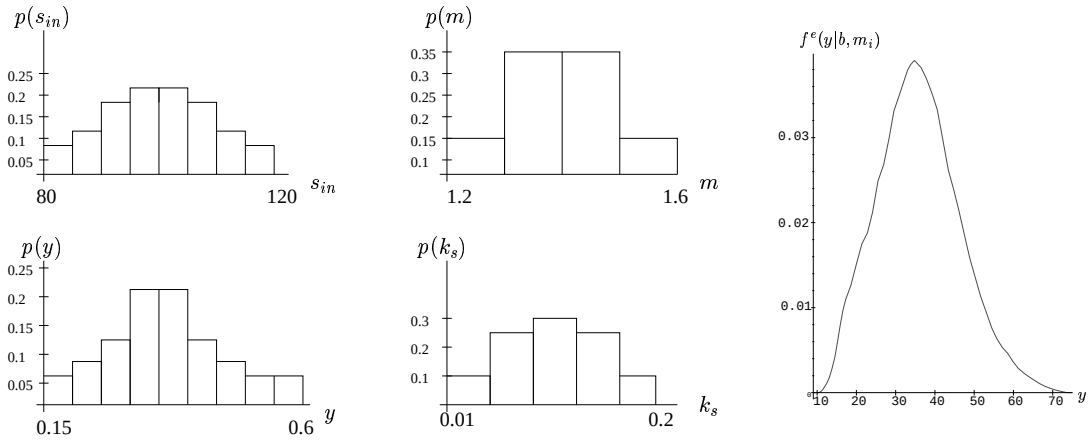


Figure 7: (a) Approximate probability distributions of the parameters y, s_{in}, m, and k_s of the Monod model, given in the form of histograms. (b) Estimation of the probability density of the value of the behavioral feature x^* (behavior b_1 of the Monod model in Fig. 5) obtained by taking into account the distributions in (a).

a uniform distribution of the mean in the predicted interval for x^* was assumed, a refined estimate of $f^e(y|b, m_i)$ is obtained (Fig. 7(b)).

Because the above approach arrives at better estimates of the model-specific pdfs, the prediction of optimal discriminatory experiments may be improved as well. However, the improved predictions come at higher computational costs. For more complex models involving a large number of parameters, these may become prohibitively large. In the example above, for instance, 1440 simulation runs were necessary to obtain the more precise estimate of (22).

The above account can be straightforwardly generalized to the case that multiple behavioral features are taken into account. Assuming the behavorial features to be independent, the corresponding joint pdf of the means is obtained by the product of the pdfs of the means of the individual behavioral features.

4.4 Properties of the expected information increment

The criterion ΔJ has some important properties which are summarized in the theorems below.[1]

The first theorem asserts that the expected information increment of any experiment is nonnegative. That is, every experiment is expected to be informative on the average. This property conforms to the intuition that the observations obtained in an experiment will help us decide between the models.

Theor. 1 The expected information increment of any experiment e is nonnegative:

$$\Delta J(e) \geq 0. \tag{24}$$

Equality holds if all models have the same predictions.

Assume that in an experiment e the value of two behavioral features p_1 and p_2 can be determined. Theor. 2 below shows that ΔJ has another desirable property: performing e and measuring the value of p_1 and p_2 is expected to be more informative than measuring the value of only one of the behavioral features.

[1] The proofs of the theorems can be found in [48].

Theor. 2 Let $\Delta J(e, p_1)$ be the expected information increment of e when p_1 is being measured and $\Delta J(e, p_1, p_2)$ the expected information increment of e when both p_1 and p_2 are taken into account. Then,

$$\Delta J(e, p_1, p_2) \geq \Delta J(e, p_1).$$

The expected information increment ΔJ, as defined here, measures the discriminatory potential of a *single experiment*. In the next section, we formulate the method for model discrimination as a sequential approach. That is, at every step the optimal discriminatory experiment is determined, this experiment is executed, and the model probabilities are updated in the light of the experimental outcome.

ΔJ can be easily adapted to predict the optimal sequence of experiments. The expected information increment of two experiments e_1 and e_2 can be estimated by computing the expected information increment of e_2 by taking into account all possible outcomes of e_1. However, determining sequences of optimal experiments for model discrimination is less efficient than determining the most informative experiment at each step and executing this experiment. In the former case, ΔJ averages the expected information increment of the second experiment according to all possible outcomes of e_1. Taking the actual outcome of e_1 will lead to a better estimate of the informativeness of e_2. In the extreme, it may happen that the results of e_1 already discriminate between the models, so that there is no need to perform e_2.

5 Algorithm for model discrimination

5.1 Description of the algorithm

On the basis of the ΔJ-criterion, a simple algorithm for the discrimination of competing models can be formulated. We assume that a model $m_i \in M$ is correct, if its probability estimate is above a predefined threshold value α, $0 < \alpha \leq 1$.

Algorithm 1 (Model discrimination) Let $p(m_i)$ be the *a priori* probabilities of the models $m_i \in M$ and E a set of possible experiments to perform.

While $\exists m_i \in M : p(m_i) \neq 0$ **and** $\forall m_i \in M : p(m_i) < \alpha$ **and** $E \neq \emptyset$ **do**

Step 1 Determine the predicted behaviors of the system for every $m_i \in M$ in the experiments $e \in E$.

Step 2 Use the model predictions to compute $\Delta J(e)$, $e \in E$. Select $e \in E$ for which $\Delta J(e)$ is maximal.

Step 3 Perform experiment e, and determine b^e, $\boldsymbol{y}^e$.

Step 4 Compute the *a posteriori* probabilities $p^e(m_i|\boldsymbol{y}^e, b^e)$ of the models. Set $p(m_i)$ to $p^e(m_i|\boldsymbol{y}^e, b^e)$.

Step 5 Remove e from E.

The algorithm iterates until a model has a sufficiently high probability ($p(m_i) \geq \alpha$), all models have zero probabilities, or all possible experiments have been executed. Obviously, if the algorithm terminates with $p(m_i) = 0$ for all m_i, the assumption that M contains a correct model is known to be violated.

In the first step of the algorithm, the competing models $m_i \in M$ are simulated by means of QSIM, Q2, and Q3 (Sec. 3) in order to derive the semi-quantitative behaviors of the system for every possible experiment $e \in E$. In step 2, the value of the expected information increment ΔJ is computed for every experiment $e \in E$. The computation of $\Delta J(e)$ exploits the predictions of the system behavior in e obtained by means of the models in M. More specifically, the predictions are used

$$(m_1) \quad \dot{x} = v \qquad \dot{v} = -\frac{k}{m}x - \frac{c}{m}v$$

$$(m_2) \quad \dot{x} = v \qquad \dot{v} = -\frac{k}{m}\left(x - \frac{x^3}{p}\right) - \frac{c}{m}v$$

$$(m_3) \quad \dot{x} = v \qquad \dot{v} = -\frac{k}{m}x - \frac{c}{m}|v|v$$

$$(m_4) \quad \dot{x} = v \qquad \dot{v} = -\frac{k}{m}\left(x - \frac{x^3}{p}\right) - \frac{c}{m}|v|v$$

$$(m_5) \quad \dot{x} = v \qquad \dot{v} = -\frac{k}{m}\left(x + \frac{x^3}{p}\right) - \frac{c}{m}v$$

$$(m_6) \quad \dot{x} = v \qquad \dot{v} = -\frac{k}{m}\left(x + \frac{x^3}{p}\right) - \frac{c}{m}|v|v$$

Figure 8: Competing models of a mass-spring system. The variables refer to the position x, velocity v, mass m, spring constant k and friction constant c. SQDEs are obtained by augmenting the models with the following intervals for the parameters: $m \in [2.95, 3.05]$, $k \in [4.97, 5.02]$, $c \in [0.3, 0.4]$, $p \in [6, 6]$. The initial range for the position is $[0.9, 1.1]$ and for the velocity $[0, 0]$. The above values are defaults, in the sense that they may be overruled in the experiments $e_1, \ldots, e_5$, as explained in the text.

to determine estimations of the conditional behavior probabilities and the model-specific probability densities (Sec. 4.2 and 4.3). The experiment yielding the highest value of the expected information increment is performed in step 3. The results of the experiment, namely the observed qualitative behavior b^e and the midpoints $\boldsymbol{y}^e$ of the values of the behavioral features, are used in step 4 to compute the posterior model probabilities via Bayes' rule. In the last step, the performed experiment is removed from the set of possible experiments E.

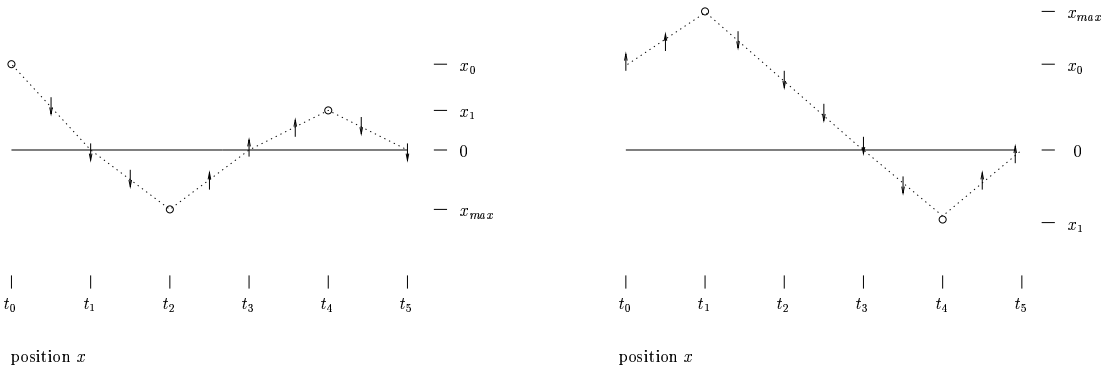
In the formulation of the algorithm, we have assumed that the experiments in E have equal costs. By selecting the most informative experiment at each iteration, the algorithm attempts to minimize the number of experiments needed to discriminate between the models. In practice, however, it may happen that different experiments have different costs. In such cases, the most informative experiment may turn out to be quite expensive. Instead of performing a single, but costly optimal discriminatory experiment, one may want to carry out several less informative, but cheaper experiments. The criterion ΔJ can then be modified to balance the informativeness of an experiment against its cost:

$$\Delta J_{cost}(e) = \frac{\Delta J(e)}{s(cost(e))},$$

where s is a monotonically increasing function of the cost of e. The choice of s will depend on the intended application: one may be interested in effective experiments without caring about expenses, or prefer less informative but also less costly tests.

5.2 Example

In this section, we will illustrate the model discrimination algorithm by means of a simple textbook example. Consider six competing SQDE models of a mass-spring system, corresponding to the ODE models given in Fig. 8. The models differ in the terms for the spring and the friction force: m_1 and m_3 assume a linear spring force, m_2 and m_4 assume a soft spring force (the stiffness of the spring decreases with the displacement), while the spring force in m_5 and m_6 is hard (the stiffness increases with the



(a) (b)

Figure 9: (a) Qualitative behavior predicted from the models m_1 to m_6 in Fig. 8 for the experiments e_1 to e_4. (b) Qualitative behavior predicted by the six models for e_5.

displacement) [46]. In m_1, m_2, and m_5 the acceleration depends linearly on the velocity. The models m_3, m_4, and m_6 assume quadratic dependency.

Suppose the following experiments can be performed:

e_1 : Investigate the system in an approximately frictionless medium ($c = 0$);

e_2 : Investigate the system in a compact medium ($c \in [2.85, 3.15]$);

e_3 : Test with an object having mass $m \in [11.95, 12.05]$;

e_4 : Test with an object having mass $m \in [0.7, 0.8]$;

e_5 : Release the object with initial velocity $v_0 \in [0.9, 1.2]$.

For each of the experiments, all models predict a single qualitative behavior for x which is the same for all models (Fig. 9). We consider two behavioral features: the maximum distance from the rest position (x_{max}), and the time-point at which the object reaches its maximum (t_{max}).

Assume for the moment that the only observed quantity is the position x, denoting the distance from the equilibrium position. What is the most informative experiment to perform? Table 1 summarizes the interval predictions for x_{max} derived from each of the models, for each of the experiments. The experiment e_1 gives rise to identical predictions from all models. It is evident, even without looking at the value of $\Delta J(e_1)$, that the corresponding experiment will not be able to distinguish between the models. The other experiments do not give distinct intervals for this feature either, but the predictions are not entirely overlapping. Hence, the measurements in the corresponding experiments may discriminate between at least some of the models.

Suppose the models have equal *a priori* probabilities $p(m_1) = \ldots = p(m_6) = 1/6$, and $\alpha = 0.85$, that is, a model is considered correct, if its probability is larger than 0.85. Furthermore, assume that the measurement of x_{max} is uniformly distributed with $\varepsilon = 0.1$. At the first step of the algorithm, e_2 is chosen since it maximizes ΔJ (see Table $2(\mathrm{a}_1)$). Assume the experiment is executed and a measurement with mean -0.108 is obtained. The measurement is not consistent with the predictions derived from m_4, m_5, and m_6 for this experiment and, hence, the *a posteriori* probabilities of these models become 0. The *a posteriori* probabilities of the other three models after the experiment are $p(m_1) = 0.029$, $p(m_2) = 0.802$ and $p(m_3) = 0.169$. In the next iteration, e_4 is selected (Table $2(\mathrm{a}_2)$). A measurement with mean -0.391 causes $p(m_1) = 0.003$, $p(m_2) = 0.983$, and $p(m_3) = 0.014$ and the algorithm terminates, returning m_2 as the correct model of the system.

x_{max}	e_1	e_2	e_3	e_4	e_5
m_1	[-1.1, -0.9]	[-0.417, -0.156]	[-1.1, -0.742]	[-1.030, -0.435]	[1.085, 1.375]
m_2	[-1.1, -0.9]	[-0.384, 0.104]	[-1.1, -0.677]	[-1.061, -0.264]	[1.150, 1.560]
m_3	[-1.1, -0.9]	[-0.382, -0.148]	[-1.098, -0.752]	[-0.993, -0.436]	[1.109, 1.436]
m_4	[-1.1, -0.9]	[-0.486, -0.320]	[-1.1, -0.789]	[-0.799, -0.372]	[1.112, 1.436]
m_5	[-1.1, -0.9]	[-0.497, -0.292]	[-1.1, -0.714]	[-0.837, -0.301]	[1.156, 1.563]
m_6	[-1.1, -0.9]	[-0.500, -0.319]	[-1.1, -0.775]	[-0.822, -0.362]	[1.088, 1.375]

Table 1: Interval predictions for the behavioral feature x_{max} derived from the six competing models for the five experiments.

ΔJ	e_1	e_2	e_3	e_4	e_5
(a_1)	0.0	0.442	0.117	0.273	0.276
(a_2)	0.0	-	0.0772	0.133	0.0984
ΔJ	e_1	e_2	e_3	e_4	e_5
(b)	0.601	0.672	0.724	0.477	0.7735

Table 2: Values for ΔJ computed for each of the experiments. (a_1) gives the value of ΔJ for experiments e_1 to e_5 in the first iteration of the algorithm for model discrimination. (a_2) gives the value of ΔJ in the second iteration of the algorithm. In (a_1) and (a_2), ΔJ is computed for the case that only x_{max} is taken into account. In (b), the value of ΔJ is computed when both x and t are observed.

Consider now the case that both the position x and the time t are observed, with $\varepsilon = 0.1$ for x_{max}, and $\varepsilon = 0.05$ for t_{max}. In this case, we find that the optimal discriminatory experiment is e_5 (Table 2(b)). Assume e_5 is performed, and the means for x_{max}, and t_{max} are 1.45, and 2.07, respectively. These values identify m_2 as the correct model with an *a posteriori* probability estimate 1. Notice that with two behavorial features, we need only one experiment to discriminate the models.

The above results have been obtained under the assumption that x_{max} is uniformly distributed in the predicted interval. We can refine the analysis by taking into account the probability distributions of the values of the initial conditions and the parameters of the models, as explained in Sec. 4.2. Assume that only a single behavorial feature, x_{max}, is taken into account.

We have divided the intervals bounding the value of m, k, and c into 2 subintervals, the interval bounding the initial value x_0 into 4 subintervals, and the interval bounding the initial value v_0 (in experiment e_5) into 3 subintervals. As a consequence, we have performed 96 simulations in e_1 (16 for each model), 192 simulations in e_1, e_3, and e_4 (32 for each model), and 576 simulations in e_5 (96 for each model). That is, a total number of 1248 simulations, whereas under the uniform distribution assumption only 30 simulations were needed. This illustrates the point that a strife at increased precision comes at the price of higher computational costs.

However, the better estimates of the pdfs for x_{max} lead to more efficient discrimination in this case. As shown in Table 3, experiment e_2 has again the highest information increment. Suppose we perform this experiment and obtain, as above, a measurement of x_{max} with mean -0.108. The measurement is not consistent with the predictions derived from m_4, m_5, and m_6 for this experiment and, hence, the *a posteriori* probabilities of these models become 0. The *a posteriori* probabilities of the other three models after the experiment are $p(m_1) = 2.811 \cdot 10^{-4}$, $p(m_2) = 0.989$ and $p(m_3) = 0.0106$. Consequently, we find in a single step that m_2 is the correct model of the system, whereas in the case of the crude estimates of the pdfs we needed an additional experiment to arrive at this conclusion.

	e_1	e_2	e_3	e_4	e_5
ΔJ	0.0	0.747	0.151	0.351	0.369

Table 3: Values for ΔJ computed for each of the experiments with the better estimates of the pdfs for x_{max}.

5.3 Implementation and performance

The algorithm for model discrimination has been implemented on a Sun Sparc-Station5 running SunOS 5.5.1. The program has been written in Common Lisp and comprises approximately 1200 lines of code.[2] The core of the implementation is the module for computing the expected information increment of an experiment. It also includes modules for determining the model-specific probability densitity functions from the model predictions, and for computing the *a posteriori* model probabilities via Bayes' rule. The program takes as input the behavior predictions of the competing models in the possible experiments, as generated by QSIM, Q2, and Q3, as well as a list of initial model probabilities. A teletype interface allows the user to specify the values of the probability threshold α and the type of distribution of the measurements of the behavioral features. The implementation has been used in the example of this section and the application described in Sec. 6.

In order to evaluate the performance of the method for model discrimination, we have investigated whether the selection of experiments according to the expected information increment criterion leads to efficient model discrimination. More specifically, we have compared the efficiency of the method by comparing the number of steps performed by the algorithm when experiments are selected according to their expected information increment with the number of steps when experiments are selected at random. For this purpose, the following strategy has been adopted [48, 51]. First, one of the competing models has been arbitrarily selected. 'Experimental' data has then been produced by randomly choosing, for every parameter of the arbitrarily-selected model, a number in the interval range of the parameter. The randomly-selected values have been used to predict exact values for the behavioral features by means of numerical simulation of the model.[3] These predicted values have been treated as the measured mean values of the behavioral features. Next, the algorithm for model discrimination has been applied a number of times using these 'measurements' for different values of ε and α.

For this performance study, we have used the competing models of the mass-spring system given in Fig. 8. The results, reported in [48, 51], have shown that, as expected, the average number of experiments performed is higher in the case of random experiment selection, independently of the value of the threshold α, and the size of the confidence interval ε. These results demonstrate the ability of the method to discriminate more efficiently between semi-quantitative models than through the random choice of experiments. The simulation study has also shown that the average number of performed experiments decreases considerably when more behavioral features are taken into account. This confirms the intuition that when more observations are made in an experiment, that is, when more information about the system behavior becomes available, better discrimination is achieved (Theor. 2). Interestingly, we also found, at least for the mass-spring example, that the use of the refined estimates of the pdfs described in Sec. 4.3 does not lead to a significant reduction of the number of experiments needed to discriminate between the models.

[2] Another, platform-independent implementation has been developed in Java.

[3] For the numerical simulations, the fourth-fifth order Runge-Kutta method implemented in Maple has been used.

6 Application: phytoplankton growth

The method for model discrimination presented in the previous section has been applied to the selection of experiments to discriminate between competing models of phytoplankton growth in a chemostat. In this section we discuss the results of the application.

6.1 Biological background

Phytoplankton is a microscopic plant living in aquatic environments. Many phytoplankton species exist, each with a characteristic size, shape, and growth properties [26]. Like terrestrial plants, phytoplankton contains chlorophyll which is necessary for photosynthesis. In photosynthesis sunlight is used as an energy source to fuse water molecules and carbon dioxide into carbohydrates (plant food). In addition to light, phytoplankton requires nutrients to grow. A nutrient compound, if not present in sufficient concentrations, can become *limiting* to the growth of phytoplankton. Nutrient limitation is often the primary factor determining the abundance of phytoplankton in any region of the world ocean. Phytoplankton plays a crucial role in marine ecosystems. Almost all animals living in the sea depend on phytoplankton for food and life-sustaining oxygen, essential to the metabolism of aerobic organisms.

As the processes regulating phytoplankton growth are difficult to study in the open sea, the growth conditions are recreated in the laboratory by means of a type of bioreactor known as *chemostat*. The chemostat is a laboratory apparatus used to study the growth properties of continuous cultures of microorganisms, such as bacteria, yeasts, or algae (Fig. 1). It has the advantage that certain of the biological parameters presumably influencing growth can be controlled by the experimenter. The chemostat is mainly used to study the growth of populations under nutrient limitation.

Chemostat experiments have been used to investigate the growth properties of the unicelullar marine alga *Dunaliella tertiolecta* under nitrate limitation. The population growth has been studied for different values of the dilution rate $D = F/V$, where F is the inflow rate and V denotes the volume of the culture vessel. After inoculation of the chemostat, the organisms undergo a stress phase in which adaptation to the new environmental conditions takes place. After this initial adaptation phase, in which $D = 0$, the value of the dilution rate is modified, and the transient behavior of the system towards its equilibrium is observed. Once the system has reached its equilibrium, a new dilution rate experiment can be performed by changing the value of D again and observing the system behavior towards the new equilibrium. Within the same run, a number of dilution rate experiments can thus be performed, each lasting a couple of weeks.

The continuous supply of medium allows one to take frequent measurements of the biomass and the concentration of the remaining nitrate in a computer-monitored environment without disturbing the evolution of the system. However, the data thus obtained are noisy, as in the case of most biological systems [9]. For instance, phytoplankton biomass is often estimated by phytoplankton biovolume, which is difficult to measure at high precision, due to the fact that *Dunaliella tertiolecta* lacks a rigid cell wall and can exhibit rapid alterations in cell volume in response to certain environmental conditions (changes in the osmotic pressure). Furthermore, the measurements of the remaining substrate concentration become highly unreliable for small concentration values, which are under the detection limit of the apparatus.

6.2 Modeling phytoplankton growth

A variety of models have been proposed that describe the growth of phytoplankton under nutrient limitation in a chemostat. All available models share the same basic idea: (a) at low (intra or extracellular) nutrient concentrations the *uptake rate* ρ (the rate of the nutrient consumption) and the *growth rate* μ are limited by, and proportional to, the nutrient concentration, whereas (b) at high nutrient concentrations the uptake and growth rates saturate and become constant [28]. In this study, four

$$\begin{aligned}\dot{x} &= \mu(s)x - Dx \\ \dot{s} &= D(s_{in} - s) - \rho(s)x \\ \mu(s) &= \mu_{max}\frac{s}{s + k_s} \\ \rho(s) &= \frac{1}{Y}\mu(s)\end{aligned}$$

(Monod)

$$\begin{aligned}\dot{x} &= \mu(s)x - Dx \\ \dot{s} &= D(s_{in} - s) - \rho(s)x \\ \mu(s,x) &= \mu_{max}\frac{s}{s + k_x x} \\ \rho(s,x) &= \frac{1}{Y}\mu(s,x)\end{aligned}$$

(Contois)

$$\begin{aligned}\dot{x} &= \mu(q)x - Dx \\ \dot{q} &= \rho(s) - \mu(q)q \\ \dot{s} &= D(s_{in} - s) - \rho(s)x \\ \mu(q) &= \bar{\mu}(1 - \frac{k_q}{q}) \\ \rho(s) &= \rho_{max}\frac{s}{s + k_s}\end{aligned}$$

(Droop)

$$\begin{aligned}\dot{x} &= \mu(q)x - Dx \\ \dot{q} &= \rho(s) - \mu(q)q \\ \dot{s} &= D(s_{in} - s) - \rho(s)x \\ \mu(q) &= \mu_{max}\frac{q - k_q}{q - k_q + k_0} \\ \rho(s) &= \rho_{max}\frac{s}{s + k_s}\end{aligned}$$

(Caperon-Meyer)

Parameter	Unit	Meaning	Interval value
D	1/day	dilution rate	-
s_{in}	μmol/l	input nutrient concentration	[80, 120]
μ_{max}	1/day	maximum growth rate	$[1.20, 1.60]$
k_s	μmol/l	half-saturation constant	$[0.01, 0.20]$
Y	$\mu m^3/\mu$mol	growth yield	$[0.15, 0.60]$
k_x	μmol/μm^3	half-saturation constant	$[0.00, 0.02]$
$\bar{\mu}$	1/day	maximum growth rate	$[1.70, 2.30]$
k_q	μmol/μm^3	minimum cell quota	$[1.60, 2.00]$
ρ_{max}	μmol/(μm^3 day)	maximum uptake rate of nutrients	$[9.25, 9.55]$
k_0	μmol/μm^3	half-saturation constant	$[2.00, 2.40]$

Figure 10: The Monod, Contois, Droop, and Caperon-Meyer models of nutrient-limited phytoplankton growth in the chemostat. The models contain the variables x, s, and q, where x [μm^3/l] denotes the total amount of biomass per unit volume, s [μmol/l] the concentration of remaining nutrient, q [μmol/μm^3] the internal cell quota. The parameters and their interval value are given in the accompanying table. The values of D and s_{in} depend on the experiment under consideration.

models for the growth of phytoplankton are considered. The models make different assumptions about the consumption of nutrients, the influence of the biomass on the growth rate of the population, and the relation between growth and uptake rates.

In addition to the Monod and Droop models presented in the introductory section, we consider two other models proposed by Contois [14] and Caperon and Meyer [13] (Fig. 10). The Contois model is a variation of the Monod model in which the growth rate of the population (μ) is assumed to be inhibited by the amount of biomass x. Similarly to the Droop model, the Caperon-Meyer model assumes intracellular storage of nutrients. However, the model differs from the Droop model by assuming a different expression for the growth rate of the population.[4]

Because of coarse and noisy data, precise numerical estimations of the values of the parameters cannot be obtained. This motivates the use of semi-quantitative models. The interval values for the parameters required in the SQDEs have been estimated from data obtained in preliminary experiments, conducted at saturating light intensities and at 18^0 C [6]. The data have been obtained in experiments carried out at the Laboratoire d'Océanographie (UMR 7093) at Villefranche-sur-Mer (France).

6.3 Experiment selection approach

We will investigate how the method for model discrimination performs when applied to a real problem: the selection of experiments for the discrimination of competing models of the growth behavior of phytoplankton in a chemostat. It will be shown that the method is applicable in the case of rather complex nonlinear models with large intervals bounding the unknown parameter values.

In order to discriminate between the four competing models presented in the previous section, we have considered chemostat experiments for the microalgae *Dunaliella tertiolecta* (Bernard and Sciandra, unpublished data). The experiments consist of changes in the value of the dilution rate D, in response to which the population attains a new equilibrium value (Sec. 6.1). We have considered 21 experiments, corresponding to equispaced values of D in the range $[0, 1]$: $D = 0, D = 0.05, \ldots, D = 1$. Taking into account 5% measurement uncertainty, the values of D become intervals. During the experiments, the biomass x and the concentration of the remaining nutrient s can be measured. This has motivated the choice of the behavioral features considered here: minima, maxima and equilibria of x (x_{min}, x_{max}, and x^*), as well as minima, maxima and equilibria of s (s_{min}, s_{max}, and s^*), if the value of these features is above the detection limit of the apparatus.

In order to obtain the predicted values of the behavioral features required for computing the expected information increment of the experiments, the models have been simulated using the qualitative simulator QSIM and its semi-quantitative extensions Q2 and Q3. In general, semi-quantitative simulation leads to the prediction of multiple behaviors for each of the models. This is a consequence of the complexity of the models and the large intervals for the parameter values and the initial conditions. We have verified that none of the qualitative behavior predictions is spurious, by comparing the output of QSIM, Q2, and Q3 with the analysis of the models by Bernard and Gouzé [8].

In the next sections, the simulation results for three consecutive dilution rate experiments are used to order the experiments according to their expected information increment. The predicted optimal discriminatory experiment is compared with the experiment that has been actually carried out. Next, data from the latter experiment is taken into account to update the probability estimates of the models and to refine the intervals for the model parameters. The equilibrium data are also used as initial conditions to simulate the models again and predict the next optimal dilution rate experiment. The model-specific probability density functions needed for the calculation of ΔJ are given by (23).

[4]Notice that, if $k_0 = k_q$, then the Caperon-Meyer model reduces to the Droop model.

6.4 Selection of initial experiment

The first experiment, performed by default, consists of an initial phase in which the cells adapt to the new experimental conditions at $D = 0$. It could be asked whether more discriminating experimental results would be obtained by setting D to another value. We have simulated the four models with QSIM, Q2, and Q3 for each of the dilution rate experiments. The following initial conditions have been used:

$$x_0 \in [0.088, 0.165],\ s_0 \in [50, 55],\ q_0 \in [5.4, 7.6].$$

These values have been determined from measurements of the biomass, the substrate concentration, and the internal quota at the beginning of the experiment.

For $D = 0$, the models predict a single qualitative behavior in which x asymptotically increases towards its equilibrium x^*, and s asymptotically decreases towards s^* (behavior b_1 in Fig. 11). For all other values of D, the four models predict a qualitative behavior in which the equilibrium of the biomass x is reached asymptotically, while the remaining substrate concentration s first reaches a maximum before decreasing towards its equilibrium (behavior b_2 in Fig. 11). Because the models predict a single behavior, either b_1 or b_2, depending on the value of D, we have $p(b_1) = 1$ and $p(b_2) = 1$.

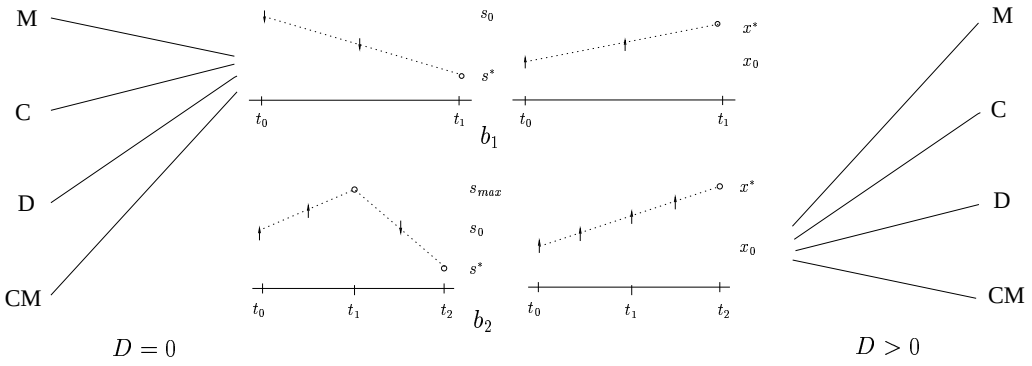


Figure 11: Selection of initial experiment. Qualitative behaviors for the growth of *Dunaliella tertiolecta* predicted by the Monod (M), Contois (C), Droop (D), and Caperon-Meyer (CM) models for all dilution rate experiments, for the initial conditions $x_0 \in [0.088, 0.165], s_0 \in [50, 55], q_0 \in [5.4, 7.6]$. b_1 is predicted by the four models for the experiment $D = 0$, while b_2 is predicted for the other dilution rate experiments.

The four models have been assumed equiprobable *a priori*, that is, $p(\mathrm{M}) = p(\mathrm{C}) = p(\mathrm{D}) = p(\mathrm{CM}) = 0.25$. Using these initial probability estimates, together with the interval predictions for the behavioral features x^*, s^* and s_{max}, as well as the calculated behavior probabilities, the value of the expected information increment ΔJ has been computed for each of the dilution rate experiments (Fig. 12).

The figure shows that the optimal discriminatory experiment is $D = 1$, with expected information increment $\Delta J = 0.84$. The predictions of the four models for the behavior of the system in this experiment are shown in Fig. 13(a). As explained above, a single qualitative behavior b_2 is predicted. The figure shows only the predictions for x^* and s^*, because the predictions for s_{max} are broad and largely overlapping for the four models. As can be seen, the predictions of the Droop and Caperon-Meyer models, although included in those of the Monod and Contois models, are much more precise. The optimal discriminatory experiment should be compared with the experiment actually performed, $D = 0$, which turns out to be the experiment with the lowest expected information increment ($\Delta J = 0.33$). Fig. 13(b) displays the predictions for x^* and s^* in $D = 0$. Notice that all four models predict $s^* = 0$, which explains the low ΔJ-value of this experiment in comparison with the experiment $D = 1$.

We have taken into account the observations made in the experiment $D = 0$ to update the probability estimates of the models. As can be seen in Fig. 14(a), x and s asymptotically reach their

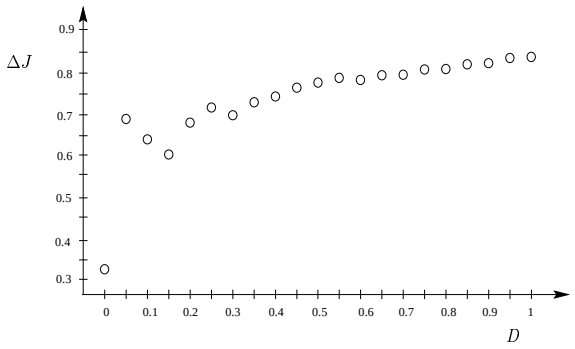


Figure 12: Selection of initial experiment. Values for the expected information increment ΔJ for each of the dilution rate experiments. $D = 1$ is the predicted optimal discriminatory experiment. Experiment $D = 0$ has been performed.

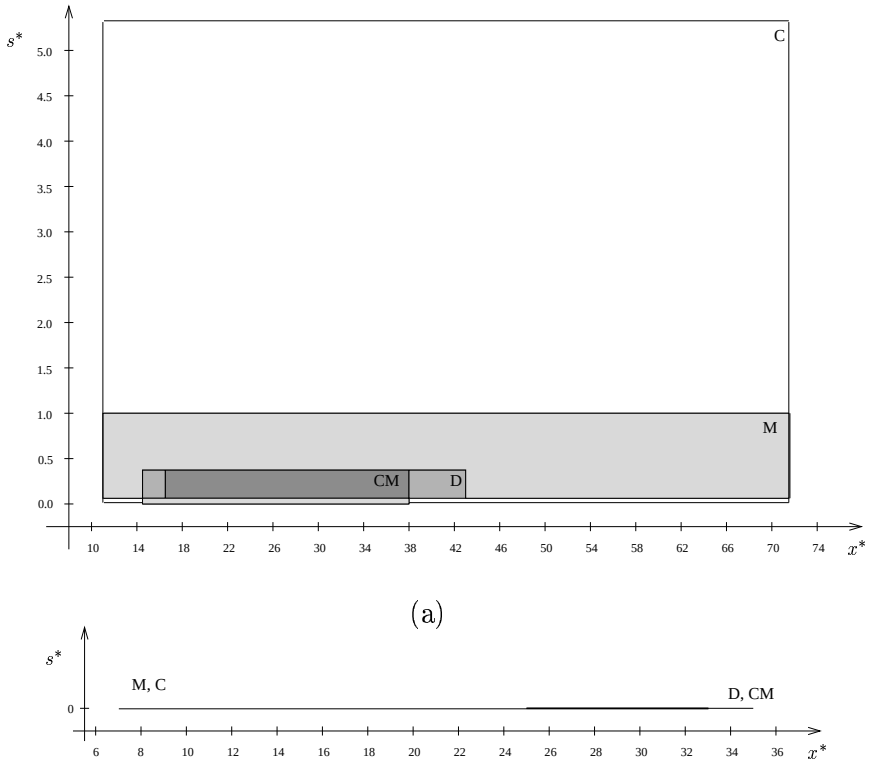


Figure 13: Selection of initial experiment. Interval predictions of the four competing models for the behavioral features x^* (behavior b_1) and s^*, x^* (behavior b_2). In (a), the predictions for the optimal discriminatory experiment $D = 1$ are given, while (b) shows the predictions for the experiment that has been actually carried out, $D = 0$. In the figures stronger shading designates more overlap between the predictions.

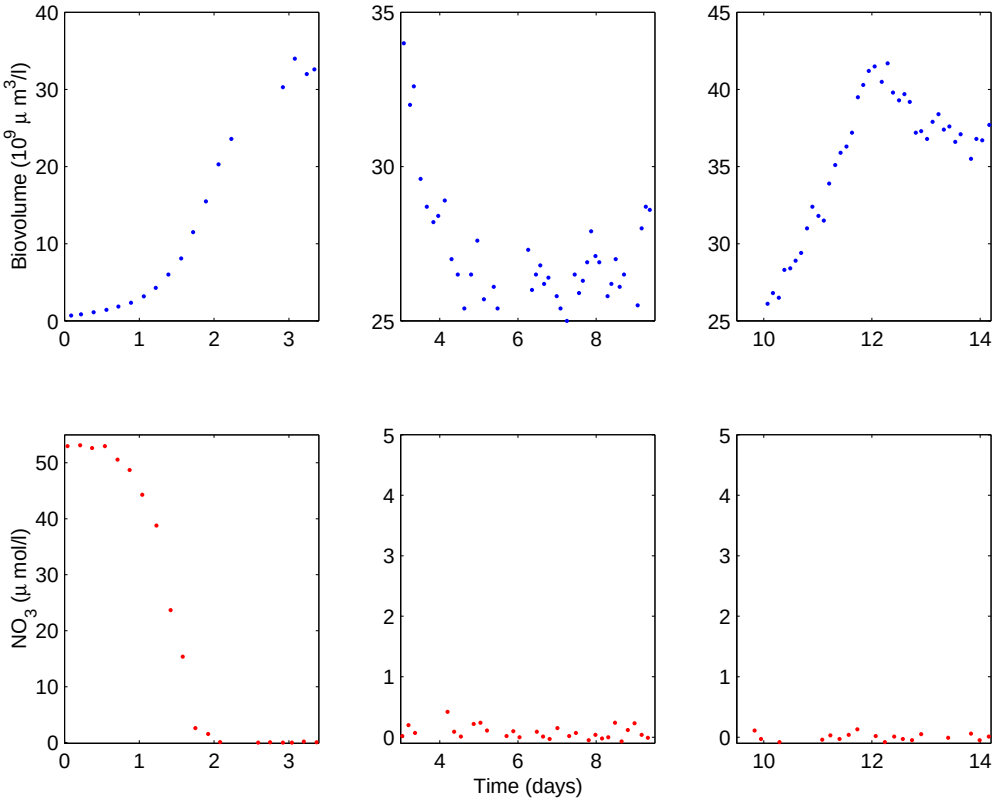


Figure 14: Plot of the measurements of the temporal evolution of the biomass x and the substrate concentration s in (a) the initial experiment ($D = 0$), (b) the second experiment ($D = 0.95$), and (c) the third experiment ($D = 0.45$) (Bernard and Sciandra, unpublished data).

equilibrium, in agreement with the predicted qualitative behavior b_1. The measured value of the behavioral feature x^* is $[30.7, 36.2]$, while $s^* \in [0, 0.01]$ is below the detection limit of the apparatus. The measurements are in agreement with the predictions of all models. As a result, none of the models can be ruled out. Using Bayes' rule the *a posteriori* probability estimates of the models can be computed:

$$p(\mathrm{M}) = 0.08,\ p(\mathrm{C}) = 0.08,\ p(\mathrm{D}) = 0.41,\ p(\mathrm{CM}) = 0.41. \tag{25}$$

The *a posteriori* model probabilities give an estimation of both the quality of the fit between the model predictions and the observations, and the precision of these predictions. The predictions of the Droop and Caperon-Meyer models are included in the observations, whereas this is not the case for the predictions of the Monod and Contois models, which explains their lower probability.

The measurements allow the parameter values in the competing models to be refined by means of the interval constraint propagation algorithms in Q2 and SQCA (Sec. 3). In this case, the measured value for x^* has been propagated through the constraint networks obtained by simulation of the actually-performed experiment $D = 0$ for each of the four models. The measurements refine the interval for the yield constant Y in the Monod and Contois model ($[0.55, 0.60]$ instead of $[0.15, 0.60]$). Furthermore, the interval for k_q in the Droop and Caperon-Meyer model is slightly narrowed from $[1.60, 2.00]$ to $[1.60, 1.96]$. These refined parameter values are used in the determination of the next optimal discriminatory experiment.

6.5 Selection of second experiment

After the system has reached equilibrium, the value of D can be changed, and the behavior of the system towards its new equilibrium observed. In order to determine which dilution rate experiment would be optimal, we have again applied QSIM, Q2, and Q3 to simulate the four competing models for each possible experiment. The new initial conditions have been determined from the equilibrium values of x, s, and q in the experiment $D = 0$. While the value of x^* has been measured in the experiment ($x^* \in [30.7, 36.2]$), the value of s^* is below the detection limit of the apparatus and set to $[0, 0.01]$. The value of q^* has not been determined experimentally. However, using an auxiliary model based on mass conservation [10], it can be inferred that $q^* \in [1.39, 1.83]$ (not shown).

Obviously, for $D = 0$ the state of the system does not change, since we start from the equilibrium reached in the previous step. For every other experiment, semi-quantitative simulation results in a single qualitative behavior for all four models (Fig. 15). The behavior consists of a minimum of x, followed by a maximum of s. Of course, the probability of this behavior, $p(b_1)$, equals 1.

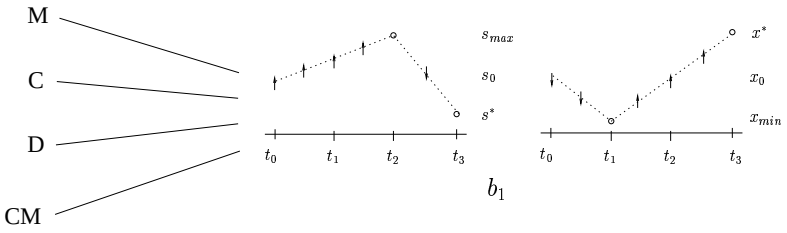


Figure 15: Selection of second experiment. Qualitative behaviors for the growth of *Dunaliella tertiolecta* predicted by the four competing models for every dilution rate experiment. For all D, $0 < D \leq 1$, a single qualitative behavior b_1 is predicted.

Using the model probabilities (25) and the simulation results, the expected information increment of each of the dilution rate experiments has been computed (Fig. 16). The results show that the experiment $D = 1$ is optimal with $\Delta J = 0.69$. Fig. 17(a) summarizes the model predictions for the features x^* and s^* (the features x_{min} and s_{max} have been omitted, because they are not very discriminatory). As can be seen, the predictions of the Monod and Contois models on the one hand, and the Droop and Caperon-Meyer models on the other, do not overlap. Even when taking into account uncertainty in the measurement of x, the performance of the experiment $D = 1$ is very likely to eliminate at least two of the four models. This is not guaranteed to be the case for the other experiments!

The experiment $D = 1$ has not been performed, but data from the experiment $D = 0.95$, with almost the same ΔJ-value, was available (Fig. 14(b)). The curve agrees with the predicted qualitative behavior, while the equilibrium of x has been measured ($x^* \in [25.3, 28.3]$) and the equilibrium of s is estimated from the detection limit of the apparatus ($s^* \in [0, 0.01]$). These results yield the following *a posteriori* model probabilities:

$$p(\mathrm{M}) = 0,\ p(\mathrm{C}) = 0,\ p(\mathrm{D}) = 0.56,\ p(\mathrm{CM}) = 0.43. \tag{26}$$

Notice that the Monod and Contois model have been eliminated by the measurements, because their predictions lie outside the measured interval for x^*. The Droop and Caperon-Meyer models are approximately equally probable, which is not surprising given that their predictions are overlapping to a large extent. The intervals for k_q are slightly improved in the latter models ($k_q \in [1.65, 1.90]$ for Droop and $k_q \in [1.6, 1.74]$ for Caperon-Meyer).

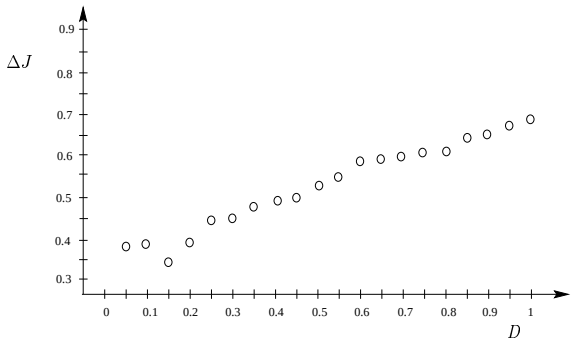


Figure 16: Selection of second experiment. Values for the expected information increment ΔJ for varying values of the dilution rate. Experiment $D = 1$ is the optimal discriminatory experiment, whereas experiment $D = 0.95$ has been performed.

6.6 Selection of third experiment

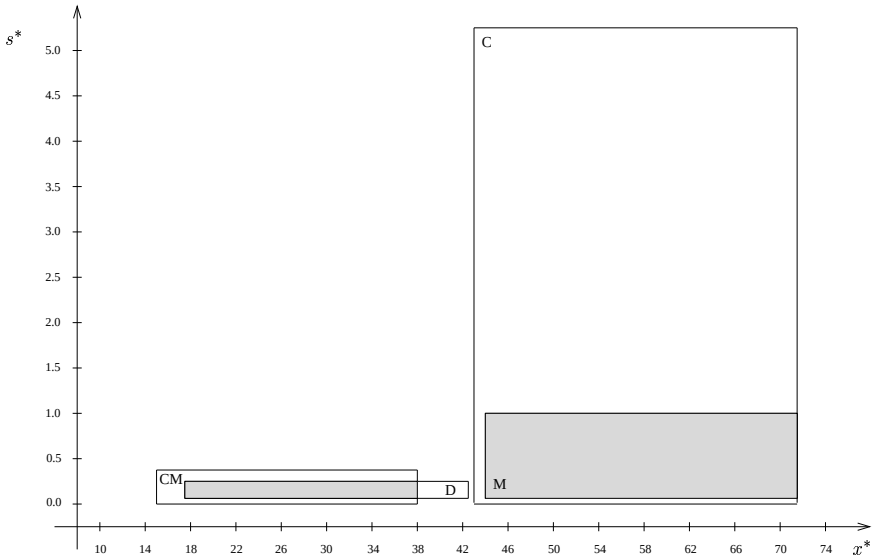
After the first two steps, we are left with only two models. Which experiment would be able to discriminate between the Droop and Caperon-Meyer model? As before, we simulate the models from the equilibrium attained in the previous step, here $x^* \in [25.3, 28.3]$, $s^* \in [0, 0.01]$, and $q^* \in [2.82, 4.74]$. For every experiment, each of the models predicts two possible qualitative behaviors for the biomass x (Fig. 18). For $D < 0.95$, the biomass is expected to increase asymptotically towards its equilibrium (behavior b_1), or to first pass through a maximum (b_2). For $D > 0.95$, the models predict that x either decreases asymptotically to its new equilibrium (behaviors b_3), or first passes through a minimum (b_4). Estimation of the behavior probabilities, as explained in Sec. 4.2, shows that $p(b_1) = p(b_2) = 0.5$, and $p(b_3) = p(b_4) = 0.5$ for all models and experiments.

Computation of the expected information increment (not shown) learns that $D = 0.9$ is the optimal discriminatory experiment, whereas $D = 0.45$ has been performed. In the latter experiment, x has been found to reach its equilibrium after passing through a maximum (Fig. 14(c)). As a consequence, behavior b_1 is ruled out. The values of the behavioral features x_{max} and x^* have been measured to be $[41.3, 43.7]$ and $[34.8, 38.3]$, respectively. These results are in agreement with the predictions of the two models, and yield the following *a posteriori* probabilities:

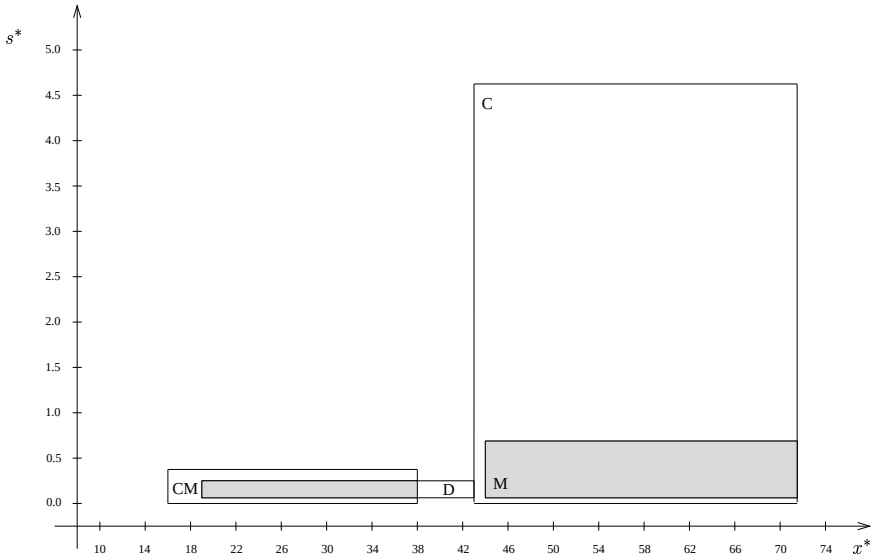
$$p(\mathrm{D}) = 0.41,\ p(\mathrm{CM}) = 0.59.$$

That is, the experiment has not succeeded in discriminating the models. The predictions of the Caperon-Meyer model are more precise than those of the Droop model (not shown), which explains the slightly higher probability of the former. We will break off the model discrimination process at this point, since no further data was available for this run of experiments. We remark just in passing that the theory would predict $D = 0.05$ as the next optimal discriminatory experiment (with $\Delta J = 0.35$).

In summary, the application illustrates that the choice of experiments plays a crucial role in model discrimination. By choosing the experiment $D = 1$ instead of $D = 0$ in the initial experiment, we might have obtained negligible *a posteriori* probabilities for the Monod and Contois model in the first step. Also, choosing an experiment with a low D-value in the second experiment would probably not have led to results inconsistent with the Monod and Contois models. This complements the results obtained in Sec. 4.4, where we showed by means of a simulation study that selecting experiments according to their expected information increment leads to more efficient discrimination than random selection of experiments. Two of the four models were ruled out in the model discrimination process, using data from experiments that were actually performed instead of predicted to be optimally discriminatory. The difficulty to discriminate between the Droop and the Caperon-Meyer models reflects the similarity of the two models. Further experiments, performed under conditions for which the predictions of the



(a)



(b)

Figure 17: Selection of second experiment. Interval predictions of the four competing models for the experiments (a) $D = 1$ (optimal discriminatory experiment) and (b) $D = 0.95$ (actually-performed experiment). The predictions concern the features x^* and s^*.

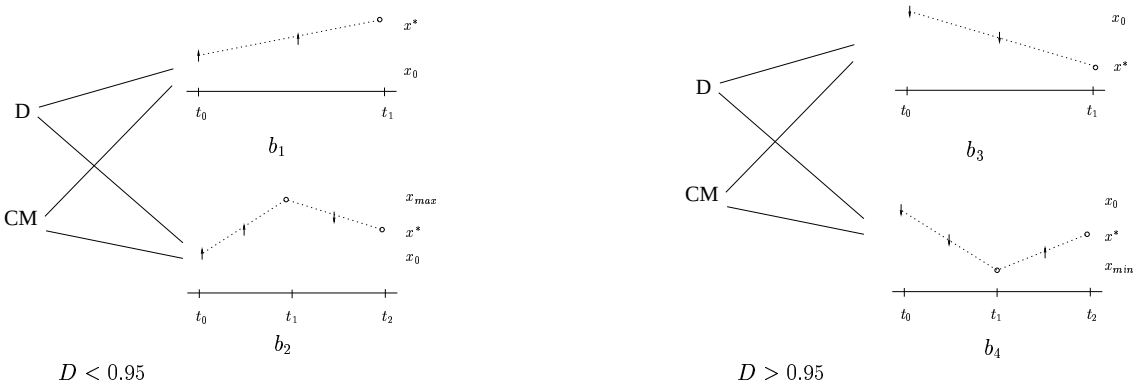


Figure 18: Selection of third experiment. Qualitative behaviors for the growth of *Dunaliella tertiolecta* predicted by the four competing models for every dilution rate experiment. Behaviors b_1 and b_2 are predicted for the experiments $D < 0.95$, while behaviors b_3 and b_4 are predicted for the experiment $D > 0.95$.

two models diverge, will be needed to achieve discrimination. Finally, the application demonstrates how, within the same reasoning framework, intervals for parameter values can be refined by using the results of conducted experiments. This allows flexible integration of experimental data in the process of model discrimination.

7 Discussion and related work

The method for model discrimination does not make any domain-specific assumptions. This guarantees that the method can be used for the discrimination between competing models of a wide range of experimental systems in a variety of scientific domains. However, there are some requirements for the method to be applicable, determined by the assumptions underlying each of the key elements of the schema of Fig. 3: the models, the experiments, the simulation techniques and the model discrimination criterion. These assumptions will be discussed below, in the context of related work.

7.1 Incompletely-specified models

In many situations, the formulation of a quantitative model is impeded by incomplete or imprecise information about the system structure or behavior. The methods for model discrimination described in the statistical literature [2, 11, 23, 24, 29, 44] are not applicable in these situations, as they require precise numerical values for the model parameters and exact functionals between the model variables to be available. The method described in this paper is able to deal with incomplete and imprecise knowledge by allowing the user to specify numerical intervals bounding parameter values and numerical envelopes bounding unknown functional relations.

Semi-quantitative models present one way of dealing with incomplete or imprecise information. Another way to achieve this would be to use qualitative models that do not need any numerical information at all. We will here describe two examples of model discrimination based on purely qualitative models, derived from different application domains.

Methods for systematically discriminating incompletely specified models have been proposed for the *analysis of genetic networks*, the networks of genes and their mutual regulatory interactions underlying the functioning of cells. Recently, a variety of experimental techniques have been developed that allow the expression level of all genes of an organism to be observed simultaneously [38, 43]. The availability of these techniques has sparked an interest in reconstructing genetic networks from experimental data. Algorithms generating models of genetic networks from gene expression profiles usually result in a set

of candidates consistent with the data. Further perturbations of the experimental system are then necessary to distinguish between the models.

Ideker *et al.* [30] have presented a method for discriminating between competing models of genetic networks by determining the optimal perturbation of the system. Because quantitative information on gene expression and regulatory interactions is generally not available, numerical models of genetic networks are difficult to handle. In the approach of Ideker *et al.* [30], Boolean networks are employed. In the framework of Boolean networks, the state of a gene is discretized (on/off) and Boolean functions are used to describe the regulatory interactions. As in this paper, an entropy criterion is used to determine which perturbation of the system is expected to be most informative with respect to the set of candidate models. In order to simplify the analysis, the genetic networks under study are assumed not to contain feedback loops and the candidate models to be equiprobable.

Another domain in which methods for the discrimination of incompletely-specified models have been developed is *model-based diagnosis (MBD)* [25]. Essentially, model-based diagnosis is concerned with explaining the malfunctioning of a (physical) device. MBD approaches usually proceed by comparing the observed behavior of the system with a reference behavior generated from a model of the normally-operating device. If a discrepancy is observed, possible faults (*candidate diagnoses*) of the system are hypothesized. The candidate diagnoses and the model of the system can be used to construct possible models of the faulty device. In order to determine which of the models reflects the actual situation, additional experiments need to be performed.

In many situations of interest, the physical devices considered in MBD can be described by models with variables taking a value from a finite, discrete domain. For instance, the voltage of a component of an electrical circuit might be assigned a value from the domain {low, medium, high}. In order to discriminate candidate diagnoses generated from such models, de Kleer and Williams [18, 19] have proposed the *General Diagnostic Engine (GDE)* (see [39, 42] for alternative approaches). The GDE method computes the *a priori* probability for each candidate diagnosis on the basis of given component failure probabilities. As in this paper, probabilities of diagnoses are updated via Bayes' rule as observations accumulate. To determine what is actually wrong with the system, additional measurements are planned. GDE follows a sequential approach by choosing the best measurement point based on an entropy criterion, performing the measurement, and then comparing the value obtained with the predictions of the candidate diagnoses. Struss [47] has extended the GDE approach by finding not only the most informative measurement point but also the most informative operating conditions of the system.

In the absence of numerical information, model discrimination approaches based on qualitative models are not able to deal with multiple predicted behaviors. For instance, both in GDE and the approach proposed by Struss, each diagnostic candidate is assumed to predict a single possible behavior of the faulty device. The same holds true for the method of Ideker *et al.*, where the models of genetic networks are also assumed to predict a single behavior after perturbation. Another restriction of the approaches based on qualitative models is their lack of discriminability. As Struss [47] remarks "... the abstract representation may be too coarse to allow for the separation of particular behaviors". Indeed, as illustrated in Sec. 6.5, two models may yield the same qualitative predictions, whereas their (semi-)quantitative predictions are different.

7.2 Criteria for experiment selection

The basic idea of model discrimination is to select the experiment that best discriminates between the competing models. We can predict the optimal discriminatory experiment by investigating the predictions of the system behavior derived from the models. Intuitively, one expects an experiment giving rise to distant predictions to be more discriminatory than an experiment for which the model predictions are close. This intuition has been formalized in statistical terms by means of concepts derived from information theory. This paper has generalized existing *maximum entropy* methods (*e.g.*,

[11, 23, 44]) to the case that the imprecision in the models does not allow exact predictions to be derived. A further generalization consists of taking into account that competing models may predict multiple qualitative behaviors.

An alternative to the maximum entropy approach is formed by the *maximum divergence* methods, which will be discussed in this section. Maximum divergence methods can be further divided into two groups: divergence in the 'average' and divergence in the 'minimum' sense. In both cases a function $F(\boldsymbol{x})$ is defined, where $\boldsymbol{x}$ is the vector of experimentally-varied quantities, and the value of $\boldsymbol{x}$ that maximizes F is sought. The methods proceed sequentially, that is, at each step a value for $\boldsymbol{x}$ is chosen, the experiment is performed, and the results are used to judge the correctness of the models.

In the case that the average divergence is maximized, the function $F(\boldsymbol{x})$ is defined as a sum of the differences of the predictions of all pairs of models. The criterion suggests that optimal discrimination will be achieved by performing the next experiment at the points where the average difference in model predictions is largest [22, 29]. A disadvantage of these methods is that they can only be applied to mathematical models with a precisely-defined structure and exact numerical values for the model parameters.

Hsiang and Reilly [27] suggest an approach to model discrimination that takes into account uncertainty in the parameter values. Parameters are assumed to have a finite number of possible values with known discrete distributions. After the optimal experiment has been performed, the observations are taken into account to update the model and parameter probabilities via Bayes' rule. The idea behind the criterion proposed by Hsiang and Reilly is closely related to the experiment selection criterion proposed in this article. It takes into account parameter uncertainties in the form of discrete probability distributions to find the expected value of the experimental outcome. However, a major difference is that, although imprecisely-specified models are permitted, the models are used to derive point predictions (the average of all possible predictions). In our method, the model predictions are intervals guaranteed to contain all possible outcomes. Another difference is that in the method of Hsiang and Reilly parameter values are allowed to take only a finite number of discrete values, whereas we allow the parameter values to be within a continuous range.

A disadvantage common to all approaches based on average divergence is that the criteria do not specify that a minimum difference in the measured variables must exist between all models. This allows the opportunity for the optimum set of inputs to discriminate a single model dramatically from the rest but not to discriminate between any of the remaining models. This problem does not occur in the maximum entropy methods or in the methods maximizing the minimum divergence between the model predictions.

In the case of maximizing the minimum divergence [2, 15], the function $F(\boldsymbol{x})$ calculates, for a given set of experimental conditions, the minimum difference between the predictions of all pairs of models. For every input $\boldsymbol{x}$, the minimum difference is calculated, and the point for which $F(\boldsymbol{x})$ reaches a maximum is chosen. Intuitively, this criterion picks the experiment for which the predictions of the closest models differ maximally.

The maximum divergence criteria employ the fact that the model predictions are real numbers. In the case of interval predictions, these algorithms cannot be used directly as the concept of distance between intervals does not have a natural definition [41]. One possibility is to employ the Kullback-Leiber divergence between two model-specific densities [36]. However, the Kullback-Leiber divergence cannot be used when the competing models predict multiple qualitative behaviors, although it might be possible to generalize this definition.

7.3 Techniques for semi-quantitative simulation

Simulation is a crucial step in the method for model discrimination. The informativeness of each experiment is evaluated on the basis of the predictions derived from the competing models. Two important requirements on the simulation techniques used to derive the predictions can be distinguished: the

techniques should derive *all possible* predictions, and the predictions must be *as precise as possible.* The requirement that all possible predictions must be derived is essential. Omitted predictions may cause a model to be falsely ruled out when the predictions derived from this model do not match the observations in an experiment. On the other hand, precise predictions are important to efficiently discriminate between the models. Predictions that overestimate the real solutions may cause a model to be corroborated, whereas it should be ruled out.

The techniques employed in this work satisfy the first requirement: QSIM, Q2, Q3, and SQCA have been proven sound [5, 34, 50]. However, due to the incompleteness of the simulation techniques, the predictions may overestimate the real solutions. As a consequence, models are only ruled out when they should be. Overestimated interval predictions lead to imprecise estimates of the model-specific probability density functions and the behavior probabilities, and hence to imprecise estimates of ΔJ. Moreover, overestimation of the intervals obtained after integration of measurements lead to imprecise estimates of the posterior models probabilities and the parameter values that are used in the next round of model discrimination.

The simulation techniques employed in this paper aim at obtaining predictions of the behavioral features that are as precise as possible. Q2 adds quantitative information to the output of the qualitative simulator QSIM, by propagating numerical bounds on parameter values and functions. Q3 may improve the results of Q2 by refining the time-step in semi-quantitative simulation. The integration of measurements can be improved through the use of SQCA, which infers additional constraints from the comparison with models describing the system in previously-performed experiments. Furthermore, by subdividing the interval ranges of variables, simulating for each of the subintervals, and combining the results, refined estimates of the model-specific probability density functions can be obtained.

It is important to bear in mind that the improvements obtained through the use of the above techniques have a cost. For example, the algorithm underlying Q3, step-size refinement, interpolates new states in a given semi-quantitative behavior, leading to an expansion of the constraint network derived from the corresponding SQDE. Consequently, constraint propagation needs more time and may strain memory resources in the case of larger models. The choice of a particular set of simulation techniques should take into account the trade-off between the precision of predictions and the computational costs. The simulation study reported in Sec. 5.3 has shown that, even for strong approximations of the model-specific probability density functions, good discriminatory experiments can be found.

7.4 Experiment selection and experiment design

The ability to discriminate between competing models of a system strongly depends upon the set of experiments that can be carried out. This set has been assumed to consist of a restricted number of experiments selected by the domain expert, taking into account constraints of feasibility and safety.

In many situations, the choice of suitable discriminatory experiments may not be straightforward to make. In the case of phytoplankton growth, for instance, the dilution rate could take values from a continuous range (Sec. 5.4). In order to deal with this, 21 equispaced values from this range have been selected. However, there is obviously no guarantee that the optimal value is included in this set.

Statistical approaches to model discrimination offer a possible solution to this problem. One can define the expected information increment of an experiment as a function of the parameters whose value is to be set in the experiments (*e.g.*, the dilution rate). The problem of finding the optimal experiment is then mapped into an optimization problem: one has to find the combination(s) of parameter values for which the function reaches its maximum [22, 23].

An even more general problem is to determine the optimal experiments when the experimental conditions can not only be constant but time-varying. In the chemostat experiments, for instance, the dilution rate can be specified by a periodic function with various possible periods and amplitudes [7]. In such cases one has to find the function that leads to efficient discrimination. The ability to deal

with these issues will shift the problem of experiment *selection* to the more general and more difficult problem of experimental *design*.

An issue related to the choice of discriminatory experiments is the selection of behavioral features. The determination of a set of behavioral features involves such decisions as when and what to observe during the experiment. We assume that the set of behavioral features consists of landmark values of the variables observed in the experiments. This is a natural choice when the predictions are in the form of semi-quantitative behaviors. However, the set of possible behavioral features is infinite when the value of an observed variable can be measured at any time-point during the experiment. To our knowledge, this problem has remained unexplored in general, although some ideas can be found in the statistical literature (*e.g.*, [3]).

8 Conclusions and further work

A method for the selection of experiments to systematically and efficiently discriminate between competing semi-quantitative models of an experimental system has been presented. The method is based on a generalized entropy criterion that ranks the possible experiments according to their discriminatory potential. The value of the criterion is computed on the basis of the model predictions obtained through semi-quantitative simulation.

The applicability of the method has been demonstrated on a real problem in the field of population biology. It has been shown that the method predicts the most informative experiments to discriminate between competing models of the growth of phytoplankton in a chemostat. This has been achieved in the presence of several complicating factors, in particular the complexity of the models, the crude estimations of the parameter values, and the difficulty to observe the behavior of the system.

The method for model discrimination, as well as the techniques employed for the derivation of the model predictions are independent of a particular application domain, as they do not employ any domain-specific knowledge about the experimental system. This guarantees that the method can be applied to the discrimination between competing models of a wide range of experimental systems. In fact, if numerical bounds on the parameters and functional relations can be formulated, and a set of discriminatory experiments exists, the method described in this paper is applicable.

Several directions for further work can be identified. First of all, the discussion in the previous section has brought to the fore the importance of simulation in the model discrimination method. In particular, overestimation of the interval bounds by semi-quantitative techniques may lead to inefficient model discrimination, in that, due to imprecise estimations of ΔH, the experiments advised by the method may not be the best experiment to perform. In order to reduce the interval bounds, the semi-quantitative simulation techniques QSIM, Q2, and Q3 have been used in this paper. However, the application of these techniques may involve high computational costs. A possible solution would be to combine semi-quantitative and Monte Carlo simulation. Semi-quantitative simulation could be used to derive guaranteed but probably overestimated intervals for the values of the model variables, and Monte Carlo simulation could be used to derive fast approximations of the probability distributions. To our knowledge, no research has been carried out in this direction.

Another promising direction for further research is the generalization of experiment selection into experiment design. This involves the ability to handle continuous ranges of parameter values and parameters given as time-varying functions. When the set of possible experiments is not given in advance, the expected information increment can be defined as a function of the unknown experimental conditions. The determination of the optimal discriminatory experiment will require the solution of a complex optimization problem. Methods for solving this optimization problem have been developed for the case of precise models with exact point predictions. Adaptation of these methods to the case of semi-quantitative models seems not to have been undertaken thus far.

More generally, the method for model discrimination described in this paper can be integrated with other tools for the computer-supported modeling of experimental systems, such as tools for automated

model building, model validation, and model revision. Model building, viewed as either the composition of a model from a set of reusable model fragments organized in a library, or as the generation of a model structure from the observed behavior of a system [33, 45, 52], may lead to a large set of alternative models of the experimental system. The selection of optimal experiments can guide the determination of the most suitable model systematically and efficiently.

Acknowledgments Most of the work described in this article was accomplished while Ivayla Vatcheva was a Ph.D. student at the Department of Computer Science, University of Twente, the Netherlands. Hidde de Jong, Nicolaas Mars, and Ivayla Vatcheva acknowledge financial support by the Programme d'Actions Intégrées franco-néerlandaises Van Gogh (dossier 99033). The authors would like to thank Jean-Luc Gouzé for his contributions to the work presented here.

References

[1] A.C. Atkinson and D.R. Cox. Planning experiments for discriminating between models. *Journal of the Royal Statistical Society*, 36:321–348, 1974.

[2] A.C. Atkinson and V.V. Fedorov. Optimal design: Experiments discriminating between several models. *Biometrika*, 62:289–303, 1975.

[3] W.G. Bardsley, R.M.W. Wood, and E.M. Melikhova. Optimal design: A computer program to study the best possible spacing of design points for model discrimination. *Computers and Chemistry*, 20(2):145–157, 1996.

[4] D. Berleant, A. Chandra, K. Bognaes, C.-G. Liaw, L. Sheng, and J. Ch'ng. Probabilities of qualitative behaviors for dependability analysis of a fault-tolerance model. In *Proceedings of the ACM/SIGAPP Symposium on Applied Computing*, pages 883–889. ACM Press, New York, NY, 1992.

[5] D. Berleant and B. Kuipers. Qualitative and quantitative simulation: Bridging the gap. *Artificial Intelligence*, 95(2):215–256, 1997.

[6] O. Bernard. *Étude Expérimentale et Théorique de la Croissance de Dunaliella tertiolecta Soumise à une Limitation Variable de Nitrate*. PhD thesis, Université Pierre & Marie-Curie, Paris, France, 1995.

[7] O. Bernard and J.-L. Gouzé. Non-linear qualitative signal processing for biological systems: Application to the algal growth in bioreactors. *Mathematical Biosciences*, 157(1-2):357–372, 1999.

[8] O. Bernard and J.-L. Gouzé. Global qualitative description of a class of nonlinear dynamical systems. *Artificial Intelligence*, 136(1):29–59, 2002.

[9] O. Bernard, G. Malara, and A. Sciandra. The effects of a controlled fluctuating nutrient environment on continuous cultures of phytoplankton monitored by computers. *Journal of Experimental Marine Biology and Ecology*, 197(2):263–278, 1996.

[10] O. Bernard, G. Sallet, and A. Sciandra. Nonlinear observers for a class of biological systems: Application to validation of a phytoplanktonic growth model. *IEEE Transactions on Automatic Control*, 43(8):1056–1065, 1998.

[11] G.E.P. Box and W.J. Hill. Discrimination among mechanistic models. *Technometrics*, 9(1):57–71, 1967.

[12] D.E. Burmaster. The unsteady continuous culture of phosphate-limited *Monochrysis lutheri Droop*: Experimental and theoretical analysis. *Journal of Experimental Marine Biology and Ecology*, 39(2):167–186, 1979.

[13] J. Caperon and J. Meyer. Nitrogen-limited growth of marine phytoplankton I: Changes in population characteristics with steady-state growth rate. *Deep-Sea Research*, 19:601–618, 1972.

[14] D.E. Contois. Kinetics of bacterial growth: Relationship between population density and specific growth rate of continuous cultures. *Journal of General Microbiology*, 21:40–50, 1959.

[15] M.J. Cooney and K.A. McDonald. Optimal dynamic experiments for bioreactor model discrimination. *Applied Microbiology and Biotechnology*, 43(5):826–837, 1995.

[16] H. de Jong. *Computer-supported Analysis of Scientific Measurements.* PhD thesis, University of Twente, Enschede, the Netherlands, 1998.

[17] H. de Jong, N. Mars, and P. van der Vet. Computer-supported resolution of measurement conflicts: A case-study in materials science. *Foundations of Science*, 4(4):427–461, 1999.

[18] J. de Kleer. Using crude probability estimates to guide diagnosis. *Artificial Intelligence*, 45(3):381–391, 1990.

[19] J. de Kleer and B.C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.

[20] M.R. Droop. Vitamin B12 and marine ecology IV: The kinetics of uptake growth and inhibition in *Monochrysis lutheri*. *Journal of the Marine Biological Association*, 48(3):689–733, 1968.

[21] D. Dvorak and B. Kuipers. Process monitoring and diagnosis, a model-based approach. *IEEE Expert*, 6(2):67–74, 1991.

[22] D. Espie and S. Macchietto. The optimal design of dynamic experiments. *AIChE*, 35(2):223–229, 1989.

[23] V.V. Fedorov. *Theory of Optimal Experiments.* Academic Press, New York, 1972.

[24] V.V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments.* Springer, New York, 1997.

[25] W. Hamscher, L. Console, and J. de Kleer, editors. *Readings in Model-Based Diagnosis.* Morgan Kaufmann, San Mateo, CA, 1992.

[26] G.P. Harris. *Phytoplankton Ecology Structure, Function and Fluctuation.* London, New York, 1986.

[27] T. Hsiang and P.M. Reilly. A practical method for discriminating among mechanistic models. *The Canadian Journal of Chemical Engineering*, 49:865–871, 1971.

[28] S.B. Hsu, S. Hubbell, and P. Waltman. A mathematical theory for single-nutrient competition in continuous cultures of microorganisms. *SIAM Journal of Applied Mathematics*, 32(2):366–383, 1977.

[29] W.G. Hunter and A.M. Reiner. Designs for discriminating between two rival models. *Technometrics*, 7:307–323, 1965.

[30] T.E. Ideker, V. Thorsson, and R.M. Karp. Discovery of regulatory interactions through perturbation: Inference and experimental design. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing, PSB 2000*, 5, pages 302–313, Singapore, 2000. World Scientific Publishing.

[31] H. Kay. SQSIM: A simulator for imprecise ODE models. *Computers and Chemical Engineering*, 23(1):27–46, 1998.

[32] H. Kay, B. Rinner, and B. Kuipers. Semi-quantitative system identification. *Artificial Intelligence*, 119(1-2):103–140, 2000.

[33] J. Keppens and Q. Shen. On compositional modelling. *Knowledge Engineering Review*, 16(2):157–200, 2001.

[34] B. Kuipers. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge, MA, 1994.

[35] B. Kuipers and D. Berleant. Using incomplete quantitative knowledge in qualitative reasoning. In *Proceedings of the 7th National Conference on Artificial Intelligence, AAAI-88*, pages 324–329. Morgan Kaufmann, Los Atlos, CA, 1988.

[36] L. Kullback and R.A. Leiber. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.

[37] R. Leitch and Q. Shen. Prioritising behaviors in qualitative simulation. In J. McDermott, editor, *Proceedings of the 13th International Joint Conference on Artificial Intelligence, IJCAI-93*, pages 1523–1528. Morgan Kaufmann, San Mateo, CA, 1993.

[38] D.J. Lockhart and E.A. Winzeler. Genomics, gene expression and DNA arrays. *Nature*, 405(6788):827–836, 2000.

[39] S. McIlraith and R. Reiter. On tests for hypothetical reasoning. In *Readings in Model-Based Diagnosis*, pages 89–96. Morgan Kaufmann, San Mateo, CA, 1992.

[40] J. Monod. *Recherches sur la croissance des cultures bactériennes*. Hermann, Paris, 1942.

[41] R.E. Moore. *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, 1979.

[42] S. Narasimhan, P.J. Mosterman, and G. Biswas. A systematic analysis of measurement selection algorithms for fault isolation in dynamic systems. In *Working Notes of the 8th International Workshop on Principles of Diagnosis, DX-98*, pages 94–101, Cape Cod, MA, 1998.

[43] A. Pandey and M. Mann. Proteomics to study genes and genomes. *Nature*, 405(6788):837–846, 2000.

[44] P.M. Reilly. Statistical methods in model discrimination. *The Canadian Journal of Chemical Engineering*, 48:168–173, 1970.

[45] C. Schut and B. Bredeweg. An overview of approaches to qualitative model construction. *Knowledge Engineering Review*, 11(1):1–25, 1996.

[46] J.J. Stoker. *Nonlinear Vibrations in Mechanical and Electrical Systems*. John Wiley & Sons, New York, 1992.

[47] P. Struss. Testing for discrimination of diagnoses. In *Working Notes of the 5th International Workshop on Principles of Diagnosis, DX-94*, New Paltz, NY, 1994.

[48] I. Vatcheva. *Computer-supported experiment selection for model discrimination*. PhD thesis, University of Twente, Enschede, the Netherlands, 2001.

[49] I. Vatcheva, O. Bernard, H. de Jong, J.-L. Gouzé, and N.J.I. Mars. Discrimination of semi-quantitative models by experiment selection: Method and application in population biology. In B. Nebel, editor, *Proceedings of the 17th International Joint Conference on Artificial Intelligence, IJCAI-01*, pages 74–79, San Mateo, CA, 2001. Morgan Kaufmann.

[50] I. Vatcheva and H. de Jong. Semi-quantitative comparative analysis. In T. Dean, editor, *Proceedings of the 16th International Joint Conference on Artificial Intelligence, IJCAI-99*, pages 1034–1040. Morgan Kaufmann, San Francisco, CA, 1999.

[51] I. Vatcheva, H. de Jong, and N.J.I. Mars. Selection of perturbation experiments for model discrimination. In W. Horn, editor, *Proceedings of the 14th European Conference on Artificial Intelligence, ECAI-2000*, pages 191–195. IOS Press, Amsterdam, 2000.

[52] S. Xia and N. Smith. Automated modelling: A discussion and review. *Knowledge Engineering Review*, 11(2):137–160, 1996.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399