Available online at www.sciencedirect.com
SCIENCE DIRECT®
Artificial Intelligence 170 (2006) 472–506
Artificial Intelligence
www.elsevier.com/locate/artint

Experiment selection for the discrimination of semi-quantitative models of dynamical systems

Ivayla Vatcheva [a], Hidde de Jong [b,*], Olivier Bernard [c], Nicolaas J.I. Mars [d]

[a] *German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany*
[b] *Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de recherche Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France*
[c] *Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de recherche Sophia Antipolis, 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France*
[d] *Materials Science Centre, Department of Mathematics and Natural Sciences, Rijksuniversiteit Groningen, Nijenborgh 4, 9747 AG Groningen, the Netherlands*

Received 19 September 2005; received in revised form 13 November 2005; accepted 13 November 2005

Available online 9 December 2005

Abstract

Modeling an experimental system often results in a number of alternative models that are all justified by the available experimental data. To discriminate among these models, additional experiments are needed. Existing methods for the selection of discriminatory experiments in statistics and in artificial intelligence are often based on an entropy criterion, the so-called information increment. A limitation of these methods is that they are not well-adapted to discriminating models of dynamical systems under conditions of limited measurability. Moreover, there are no generic procedures for computing the information increment of an experiment when the models are qualitative or semi-quantitative. This has motivated the development of a method for the selection of experiments to discriminate among semi-quantitative models of dynamical systems. The method has been implemented on top of existing implementations of the qualitative and semi-quantitative simulation techniques QSIM, Q2, and Q3. The applicability of the method to real-world problems is illustrated by means of an example in population biology: the discrimination of four competing models of the growth of phytoplankton in a bioreactor. The models have traditionally been considered equivalent for all practical purposes. Using our model discrimination approach and experimental data we show, however, that two of them are superior for describing phytoplankton growth under a wide range of experimental conditions.

Keywords: Qualitative and semi-quantitative modeling and simulation; Model discrimination; Information theory; Population biology; Computer-supported modeling

* Corresponding author.
E-mail addresses: I.Vacheva@dkfz-heidelberg.de (I. Vatcheva), Hidde.de-Jong@inrialpes.fr (H. de Jong), Olivier.Bernard@sophia.inria.fr (O. Bernard), N.J.I.Mars@fwn.rug.nl (N.J.I. Mars).

0004-3702/$ – see front matter
doi:10.1016/j.artint.2005.11.001

1. Introduction

Finding an adequate model of the functioning of a complex system—either natural or man-made—is a problem confronting practitioners in many different domains. For instance, it arises when a population biologist studies an ecosystem, an electrical engineer diagnoses a fault in an electronic circuit, or a medical doctor tries to infer the causes of the illness of a patient. Given the pre-eminence of modeling in human endeavor, it comes as no surprise that methods for computer-supported modeling have been subject to active research in artificial intelligence. For instance, machine learning techniques infer a model from observations of the system behavior [47], while automated modeling techniques in qualitative reasoning are able to derive a model from a description of the system, a user question, and a domain theory [40,51,63].

The complexity of the task of modeling is such that in many cases the available information does not allow one to decide between alternative models of the system under study. In the above examples, several interactions between species in an ecosystem can be imagined, the fault in the electronic circuit may be due to the failure of different components, and the symptoms of the patient might have a variety of causes. The different assumptions on the structure and behavior of the system may result in a number of competing models, all justifiable by the available observations. The existence of this set of competing models gives rise to the problem of *model discrimination*.

To discriminate between a number of competing models, and identify which of them most adequately describes the actual situation, new observations are needed. These can be obtained by performing additional experiments on the system. An experiment discriminates among 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 a large number of experiments can be performed, and the cost of each of them may be considerable, it is important that the experiments be selected carefully. In fact, it is desirable that experiments be selected in such a way that the set of competing models is maximally reduced (systematic model discrimination) at minimal costs (efficient model discrimination). Notice that the efficiency requirement does not necessarily amount to minimizing the number of experiments, since several cheap experiments may cost less than a single expensive experiment.

The problem of model discrimination has received substantial attention in the statistical literature (see [28,29] for reviews). This has resulted in criteria for determining the experiment that optimally discriminates among different models of the system. These criteria, based on concepts derived from information theory and optimization theory, have been applied to a variety of problems in biology and biotechnology (e.g., [15,37,55,57]), in physics (e.g., [16,45]), and in chemical engineering (e.g., [1,25]). Similar criteria have been developed in artificial intelligence, especially in the field of *model-based diagnosis* (*MBD*) [31]. In this case, the criteria play a key role in reducing the number of candidate diagnoses generated, that is, the set of hypotheses accounting for the observed faulty behavior of a device. In order to discriminate among the candidate diagnoses, and thus find out what is actually wrong with the system, methods for selecting the optimal measurements or experimental conditions have been developed (e.g., [20,22,53]).

All of the above methods share the same underlying intuition. To evaluate the discriminatory potential of an experiment, its outcome is predicted by each of the competing models. The experiment for which the model predictions differ most will have the highest chance of discriminating among the models, and is therefore selected as the optimal discriminatory experiment. Often, this intuition is formalized by means of the optimal *information increment*, that is, the maximal difference in entropy before and after the execution of an experiment. The expected value of this information increment, for each of the available experiments, can be determined from the model predictions.

A first problem with existing methods for model discrimination is that the information increment criteria being used are not well-adapted to dynamical systems. The criteria take into account the possible states of the system, but not their temporal ordering. It has been shown that under certain conditions, the temporal evolution of the state of the system is not necessary for model discrimination [53,54]. These conditions, however, are often not fulfilled in practice. As a consequence, it becomes important to take into account the differences in temporal behavior.

A second problem with existing methods is the difficulty of computing the expected value of the information increment from the model predictions. This is especially true when qualitative or semi-quantitative models are used, like those developed in the field of *qualitative reasoning* (*QR*) [26,56,62]. *Qualitative* models deal with incomplete or imprecise information by assigning symbolic values to its parameters and initial conditions [42]. In addition, qualitative models may contain incompletely specified functional relations between the variables, defined by their monotonicity properties only. Qualitative models can be extended to *semi-quantitative* models by specifying numerical intervals bounding parameter values as well as envelopes bounding monotonic functions [4,38]. The latter models are useful

in many scientific and engineering domains, which often deal with parametric and functional tolerances described by intervals and bounding curves. However, existing methods for model discrimination either focus on numerical models or, if they allow qualitative or semi-quantitative models, do not provide a generic, domain-independent computational framework for computing the expected information increment.

The above limitations of methods for model discrimination in statistics and in model-based diagnosis have motivated the work described in this paper. We present a method for the systematic and efficient discrimination of semi-quantitative models of dynamical systems [58,59]. The selection of the optimal discriminatory experiment is based on an entropy criterion that exploits the model prediction for the temporal evolution of the system state. In addition, the method provides computational procedures to actually calculate the entropy criterion from the model predictions. Overall, model discrimination proceeds iteratively, similarly to a sequential diagnosis strategy [22,30]. The algorithm starts 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 and refine the parameter intervals. 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 method has been implemented in Common Lisp, on top of the qualitative and semi-quantitative simulators QSIM, Q2, and Q3.

We will demonstrate the effectiveness of the method in the context of a real application in population biology. In this domain competing models are bound to occur due to the fact that most models are based on empirical relationships between the variables and have rarely been appropriately validated. More precisely, the problem is to discriminate between four competing models of nutrient-limited phytoplankton growth in a chemostat. The choice of optimal discriminatory experiments is critical in this application, because the experiments take several weeks to complete. Semi-quantitative models are appropriate because, as in the case of most biological systems, the available data is noisy and fragmentary. The four models have traditionally been considered equivalent for all practical purposes. By applying our model discrimination approach and employing real experimental data we show, however, that two of them are superior for reproducing phytoplankton growth under a wide range of experimental conditions. These so-called Droop and Caperon–Meyer models, which make closely related assumptions on the underlying growth processes, strike the right balance between empirical validity and mathematical simplicity.

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 (Section 3) that guides the selection of optimal discriminatory experiments. In Section 4, the algorithm for model discrimination is presented in detail. The results of the application of the method to the modeling of nutrient-limited phytoplankton growth in a chemostat are presented in Section 5. In Section 6, the method for model discrimination is discussed in the context of related work, while the last section summarizes the contributions and indicates directions for further research.

2. Basic concepts and outline

This section provides the framework for the method for model discrimination. The concepts of experimental system, experiment, observation, model, and competing models are introduced, and an outline of the method is given.

2.1. Experimental systems and experiments

An *experimental system* is a (physical, chemical, biological) system that is created and sustained in an experimental set-up [18]. The experimental set-up defines how the system is put together and the range of experimental conditions. 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. Here, we are concerned with *dynamical systems*. That is, systems whose state evolves over time.

For example, consider the experimental system consisting of 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 at a fixed rate to a culture vessel whose contents are continuously mixed. Medium, cells, and by-products are continually removed, maintaining the culture in the vessel at a constant volume. The experimenters can control, among other things, the inflow rate of

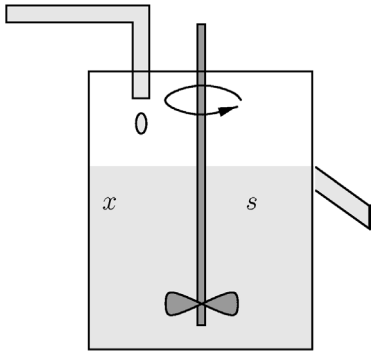


Fig. 1. Schematic illustration of the chemostat.

the growth medium, the light intensity, and the temperature. In addition, they can observe the phytoplankton biomass and the concentration of remaining nutrients over time.

An *experiment* is an action, in which specific experimental conditions are imposed, and the temporal behavior of the system is observed under these conditions. In the context of phytoplankton growth, an experiment might consist of observing the evolution of the phytoplankton biomass from given initial conditions, while the concentration of the feeding substrate is fixed.

2.2. Observations of experimental systems

The state of an experimental system evolves over time and can be described by a set of continuous variables, so-called *state variables*. More precisely, a state variable x is a function of time defined as $x : \mathbb{R}^* \to D_x$, $D_x \subseteq \mathbb{R}^*$.[1] In many cases, the state of the system cannot be completely determined because of its *limited measurability*. This means that some state variables cannot be observed over the time-course of the experiment. Usually, only certain quantities that are (continuous) functions of the state variables can be observed, here called *observed variables* and denoted by the vector $\boldsymbol{y}$. In practice, the value of $\boldsymbol{y}$ is measured at specified sampling points only and the precision of the measurements is limited. Consequently, the available measurements are often restricted to qualitative and semi-quantitative properties of the behavior of the system, as defined more formally in the remainder of this section.

Following [42], we define the domain of a variable y by its *quantity space*. The quantity space is a set of totally-ordered *landmark values* or *landmarks*, corresponding to qualitatively-significant values of the variable. Some of the landmark values are defined at the beginning, such as initial values. Others may be introduced during the temporal evolution of the system, such as extreme values and equilibrium values. Most of the time, exact landmark values are not known and we denote them by a symbol. The vector $\boldsymbol{l}$ denotes all the landmarks of the observed variables $\boldsymbol{y}$.

Time is also a variable, with a quantity space composed of landmarks of a special kind, so-called *distinguished time-points*. At each distinguished time-point t_j, at least one variable changes its value to or from a landmark. The *qualitative value* of y at a time-point t_j, or in an interval between two successive 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, or between two time-points, consists of a tuple of qualitative values, one for each variable of the system.

The *qualitative behavior* of the system on a time-interval $[t_0, t_n]$ is given by the sequence of qualitative states traversed by the system on that interval. That is, a qualitative behavior is defined by a sequence of qualitative states, alternating between states that hold at a time-point and states that last over a time-interval (see [42] for details). Each point state denotes a qualitatively significant event in the time evolution of the system, in the sense that a landmark value of at least one variable is reached.

As remarked above, landmark values are usually unknown. But it may be possible to bound the values by numerical intervals. In this case, the concept of qualitative behavior generalizes to that of *semi-quantitative behavior* [39,42,58].

[1] $\mathbb{R}^*$ represents the extended set of real numbers, including $-\infty$ and ∞.

By extension, we sometimes speak of the semi-quantitative behavior of a subset of the variables instead of that of the entire system.

An *observation* on a system in an experiment can now be formally defined.

Definition 1 *(Observation).* An observation on an experimental system in an experiment e is a semi-quantitative behavior of the observed variables $\boldsymbol{y}$ with landmark values $\boldsymbol{l}$ over the time-interval of the experiment.

In many cases it is useful to adopt a slightly modified definition of observation. For example, we can have observations in which only some of the landmarks in $\boldsymbol{l}$ are measured. In addition, landmarks that are defined as functions of the landmarks in $\boldsymbol{l}$ could be measured, such as the amplitude and the period of an oscillation. In what follows, we make the assumption that the sampling frequency of the measurements is chosen in such a way that the observed qualitative behavior of $\boldsymbol{y}$ corresponds to the real qualitative behavior of $\boldsymbol{y}$. That is, we exclude observation gaps.

In the context of phytoplankton growth, the observed variables are the amount of biomass and the concentration of remaining substrate. The observations thus consist of the semi-quantitative behavior of these variables. For instance, an observation may show that the biomass first reaches a maximum, followed by a steady state, while at the same time the substrate concentration decreases to its steady state. This gives rise to the introduction of two landmark values, corresponding to the maximum and equilibrium values of the observed variables. The observation may also provide interval bounds on these landmarks.

2.3. Models of experimental systems

The relations between the quantities of a dynamical system are traditionally modeled by ordinary differential equations (ODEs), that is, by equations of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{p}), \qquad \boldsymbol{x}(t_0) = \boldsymbol{x_0}, \qquad \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{p}) \tag{1}$$

where $\boldsymbol{x}$ is the vector of state variables of the system, $\boldsymbol{u}$ the vector of control variables, $\boldsymbol{y}$ the vector of observed variables, $\boldsymbol{x_0}$ the vector of initial conditions, and $\boldsymbol{p}$ the parameter vector. The state variables and the observed variables are functions of time. A control variable is generally a function of time as well, that is, $u : \mathbb{R}^* \to D_u$, $D_u \subseteq \mathbb{R}^*$. However, we restrict our attention here to time-invariant controls. The parameters are time-invariant by definition.

Fig. 2 shows two ODE models describing the nutrient-limited growth of phytoplankton in a chemostat. The Monod model [48] assumes that the consumed nutrient is instantaneously transformed into biomass. This assumption is expressed by the linear proportionality between the growth rate μ and the nutrient uptake rate ρ. The Droop model [11, 23] uncouples growth rate from external nutrient concentration by introducing an intracellular store of nutrients. The growth rate μ is hypothesized to depend on a quantity q called the cell quota, the average amount of stored nutrients in each cell.

In cases when knowledge on the experimental system is incomplete or imprecise, ODEs can be abstracted to *qualitative differential equations* (*QDEs*) [42]. In a first step, the ODE is decomposed into a set of basic mathematical equations, possibly introducing additional auxiliary variables besides the variables distinguished in (1). Second, all variables and parameters are assigned a quantity space with landmarks, and the basic equations are mapped to constraints between the qualitative values of the variables and parameters.

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

Fig. 2. Monod and Droop models describing the nutrient-limited growth of phytoplankton in a chemostat. x denotes the total amount of biomass per unit volume, s the concentration of remaining nutrient, q the internal cell quota. The following parameters are used: dilution rate D, input nutrient concentration s_{in}, maximum growth rate μ_{max}, half-saturation constant k_s, growth yield Y, maximum growth rate $\bar{\mu}$, minimum cell quota k_q, and maximum uptake rate of nutrients ρ_{max}. See Fig. 8 for the units of the variables and parameters in the models.

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:

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

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

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

Often, some quantitative information is available in the form of numerical ranges for parameters and initial conditions, or numerical envelopes for functional relations. In such cases, a QDE can be extended to a *semi-quantitative differential equation* (*SQDE*). For instance, the QDEs obtained from the models in Fig. 2 can be converted into SQDEs by specifying interval bounds on the parameters and initial conditions:

$$\begin{aligned} &\mu_{max} \in [1.2, 1.6], \quad k_s \in [0.01, 0.2], \quad Y \in [0.15, 0.6], \quad s_{in} \in [80, 120], \\ &\bar{\mu} \in [1.7, 2.3], \quad \rho_{max} \in [9.25, 9.55], \quad k_q \in [1.6, 2.0], \quad D \in [0.38, 0.42], \\ &x_0 \in [0.088, 0.165], \quad s_0 \in [44, 46], \quad q_0 \in [1.6, 7.62]. \end{aligned} \tag{2}$$

More precisely, in an SQDE the intervals are associated with landmark values in the quantity space of the variables and parameters. When this does not lead to ambiguities, we will often simply speak of the interval bounds on the parameters and initial conditions, as in (3). In the remainder of this article, we assume that experimental systems are modeled by semi-quantitative differential equations.

Predictions from SQDEs can be derived by means of semi-quantitative simulation. A variety of *semi-quantitative* or *interval simulation techniques* have been developed (reviewed in [17,56,58]). Here, we employ techniques that have been developed in the field of qualitative reasoning, in particular the technique QSIM [42], and its extensions Q2 [17,42,43] and Q3 [4]. Essentially, semi-quantitative simulation consists of refining the predictions of the qualitative simulator QSIM, through the integration of quantitative information by Q2 and Q3. In general, however, the model discrimination method presented in this paper does not depend on the particular set of simulation techniques. The only requirement is that the simulation techniques yield semi-quantitative behavior predictions of the observed variables.

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 them into semi-quantitative behaviors, by computing or refining numerical bounds for the landmark values. The behaviors of the observed variables, resulting from the simulation of the Monod and Droop models (Fig. 2) with interval bounds (3), are shown in Fig. 3. The table in the figure shows predicted interval bounds for the landmarks introduced during simulation.

QSIM, Q2, and Q3 have been proven *sound* [4,42,43]. That is, all possible semi-quantitative behaviors consistent with an SQDE are generated by the simulation algorithms. On the other hand, QSIM and its extensions are *incomplete*, in the sense that spurious qualitative behaviors may be generated or landmark bounds overestimated.[2] Given a predicted semi-quantitative behavior, the interval bounds on parameter values can be refined after execution of the experiment by integrating measurements of observed variables [24,39]. 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 [58,60].

2.4. Competing models of experimental systems

To compare models with observations, we need to define what it means for a model and an observation to be *consistent*. In what follows, let m denote an SQDE model. We define $\boldsymbol{y}_m$ to be the observed variables occurring in model m. Note that in general $\boldsymbol{y}_m$ will be a subset of the entire set of observed variables $\boldsymbol{y}$.

[2] The use of the terms soundness and (in)completeness is consistent with the interpretation of simulation as a logical inference process [41]. The disjunction of (qualitative or semi-quantitative) behaviors generated by simulation is true, in the sense that it includes all genuine behaviors (soundness). However, simulation may not be able to infer the true disjunction with only genuine and no spurious behaviors (incompleteness). See [52] for an alternative interpretation.

Definition 2 *(Consistency of predictions and observations).* A predicted and an observed semi-quantitative behavior of an experimental system are consistent, if $\boldsymbol{y}_m$ is not empty and

(1) The observed sequence of qualitative values for $\boldsymbol{y}_m$ is the same as the predicted sequence of qualitative values;
(2) The interval bounds on the corresponding landmark values in the predicted and observed behavior overlap.

By extension, the model m is consistent with an observed semi-quantitative behavior, if at least one of the predicted semi-quantitative behaviors is consistent with the observation.

If an SQDE model is consistent with the observed behavior in the entire range of possible experimental conditions, we say that the model is *correct*. This suggests which criteria an SQDE has to satisfy so as to qualify as a *candidate model* of the experimental system. In particular, we require that the SQDE be consistent with all available observations (which usually cover only a small fraction of the range of possible observations).

Definition 3 *(Candidate model).* An SQDE model m is a candidate model of an experimental system, if for each already performed experiment, the observed semi-quantitative behavior is consistent with the model, where consistency is defined as in Definition 2.

Consider again the example of phytoplankton growth in a chemostat. The observed variables include the amount of biomass x and the concentration of remaining nutrient s. These are state variables in the Monod model M, so $\boldsymbol{y}_{\mathrm{M}} = [x, s]^T$. If for all available experiments the predicted values for x and s are consistent with the observations, then according to Definition 3, the Monod model qualifies as a candidate model of the experimental system.

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

Definition 4 *(Competing models).* Let M be a set of candidate SQDE models. Furthermore, let $\boldsymbol{y}_M$ denote the observed variables shared by all models in the set. The models in M are competing, if

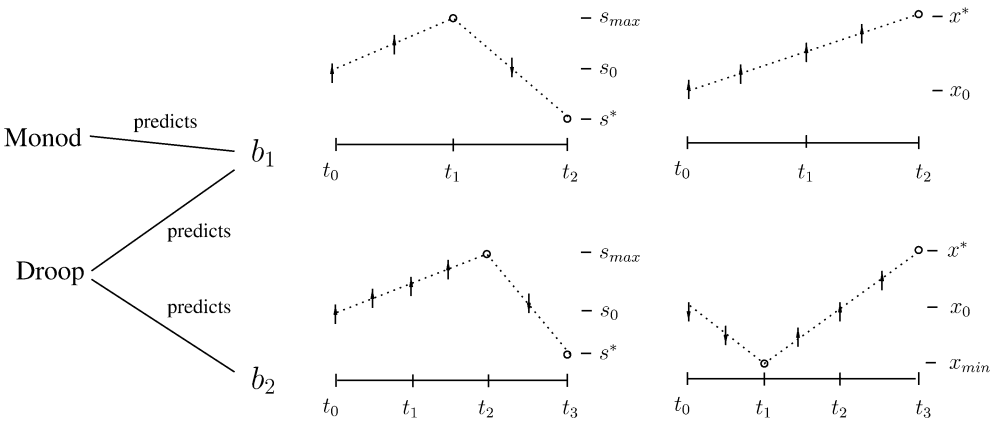
(1) $\boldsymbol{y}_M$ is not empty;
(2) For every pair of models $m, m' \in M$, one can find some experimental conditions for which the semi-quantitative behavior predictions for $\boldsymbol{y}_M$ differ.

The definition states that m and m' are competing if they predict different semi-quantitative behaviors for $\boldsymbol{y}_M$ in the same experiment. Difference here means that either the predicted sets of qualitative behaviors differ or, in case the models predict exactly the same set of qualitative behaviors, that the numerical bounds on the corresponding landmark values in at least one behavior do not completely overlap.

In what follows, we assume for notational simplicity that the observed variables $\boldsymbol{y}$ occur in all models in M, that is, $\boldsymbol{y} = \boldsymbol{y}_M$.

As can be seen from Fig. 2, the Monod and the Droop models share two observed variables, namely the amount of biomass x and the concentration of the limiting nutrient s. Furthermore, the simulation results of Fig. 3 show that for the given experimental conditions, the models predict different sets of qualitative behaviors of these variables. Hence, according to Definition 4, the two models are competing.

For each m in the set of competing models M, we define $p(m)$ to stand for the probability that m is a correct model of the experimental system in the range of experimental conditions. *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, using the Bayes' rule.



	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]	$[0.52, \infty]$	$[1.17, \infty]$	

Fig. 3. Semi-quantitative behaviors of the amount of biomass x and the remaining substrate concentration s, as predicted by the Monod and Droop models with the numerical ranges (3). The Monod model predicts a single semi-quantitative behavior b_1 in which 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 b_1 and b_2. Behavior b_1 is qualitatively equivalent to the behavior predicted by the Monod model. In behavior b_2, x first reaches a minimum, followed by a maximum of s, and then the equilibrium is approached. s_{max}, s^*, x_{min}, and x^* denote the corresponding landmark values, whose predicted interval bounds are given in the table.

2.5. Experiment selection for model discrimination

Discrimination of 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.

For each $e \in E$, every $m \in M$ has to be simulated with the appropriate parameter values and initial conditions, in order to determine the predictions of m for the semi-quantitative behavior of the system in e. 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 (Fig. 4). The expected information increment is a criterion that measures the discriminatory potential of an experiment. Intuitively, an experiment for which the model predictions differ more can be expected to lead to better discrimination, as the experimental outcome will generally agree with fewer of the predictions. In Section 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 observation thus obtained is used to adjust the probability estimates $p(m)$ of the models $m \in M$, as well as to refine the interval bounds on the parameter values. The experiment discriminates between the models in M, if the observation has changed the probability estimates of the models. Optimal discrimination will be achieved if the observation is consistent with the predictions of a single model only. If the observation is 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. 5) is repeated until one of the following happens: a model has a sufficiently high probability, all models in M have zero probability (because they are inconsistent with the observations), or the set of possible experiments E is exhausted.

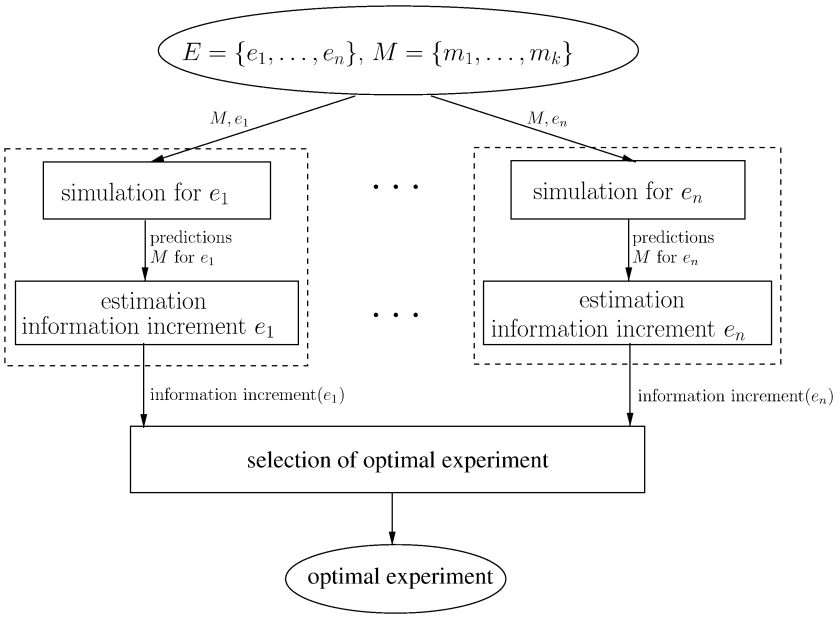


Fig. 4. Schematic overview of the selection of experiments for model discrimination.

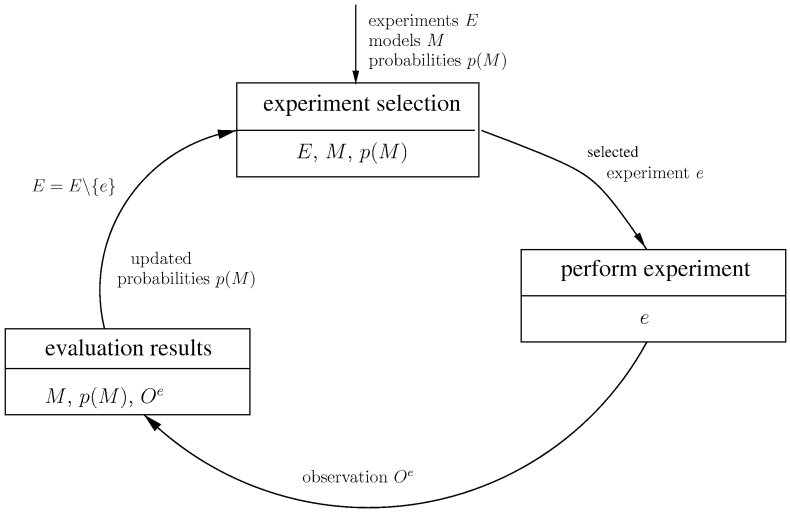


Fig. 5. Model discrimination as an iterative process.

3. Criterion for experiment selection

In this section we propose a criterion for the selection of experiments discriminating between competing semi-quantitative models [58,59]. The criterion is formalized by means of concepts from information theory and generalizes upon previous work in statistics and model-based diagnosis.

3.1. Information increment of experiment

Let M be a set of competing models of an experimental system with *a priori* probabilities $p(m)$ for every $m \in M$. We make the common assumption that M contains a single correct model, in the sense of Section 2.4:

$$\sum_{m \in M} p(m) = 1. \tag{3}$$

Although (3) is a strong assumption, the practical consequences are limited, because its violation can be tested. We further assume that a set of possible experiments E on the experimental system is given, and let O^e denote the observation made in experiment $e \in E$.

A standard measure in information theory is the *information increment* of an experiment $e \in E$ [9,28]. The information increment is defined as the difference in entropy before and after the execution of the experiment:

$$\Delta H(O^e) = -\sum_{m \in M} p(m) \ln p(m) + \sum_{m \in M} p^e(m \mid O^e) \ln p^e(m \mid O^e), \tag{4}$$

where $p^e(m \mid O^e)$ denotes the *a posteriori* probability of m, given the observation O^e in e. As can be easily verified, ΔH reaches its maximum when all posterior probabilities but one are zero. That is, when the observation made in e confirms the predictions of a single model. On the other hand, a minimal value is attained, when all posterior probabilities are equal.

Recall that in our case the observation O^e has the specific form of a semi-quantitative behavior (Definition 1). Execution of an experiment gives rise to the tuple $O^e = \langle \boldsymbol{l}^e, b^e \rangle$, where b^e denotes the observed qualitative behavior, and the vector $\boldsymbol{l}^e$ denotes measurements of the landmarks of the observed variables, for instance their minima, maxima, or equilibrium values. This allows us to write (4) as follows:

$$\Delta H(\boldsymbol{l}^e, b^e) = -\sum_{m \in M} p(m) \ln p(m) + \sum_{m \in M} p^e(m \mid \boldsymbol{l}^e, b^e) \ln p^e(m \mid \boldsymbol{l}^e, b^e), \tag{5}$$

with $p^e(m \mid \boldsymbol{l}^e, b^e)$ the *a posteriori* probability of m, given b^e is the qualitative behavior observed in e, and $\boldsymbol{l}^e$ the vector of measurements of landmark values.

Criteria for evaluating the discriminatory potential of an experiment have also been proposed in statistics (e.g., [2, 25,28,29,35]) and in model-based diagnosis (MBD) (e.g., [20,22,53]). But they are less general than the information increment (5) proposed here.

In particular, work in statistics omits the qualitative behavior of the system, which means that (5) reduces to

$$\Delta H(\boldsymbol{l}^e) = -\sum_{m \in M} p(m) \ln p(m) + \sum_{m \in M} p^e(m \mid \boldsymbol{l}^e) \ln p^e(m \mid \boldsymbol{l}^e). \tag{6}$$

The most sophisticated criteria in MBD [53] take into account some qualitative features of the system dynamics, in that they assume the value of a landmark to be measured while the system is in a certain qualitative state. This gives rise to the following expression of the information increment of an experiment:

$$\Delta H(\boldsymbol{l}^e, S^e) = -\sum_{m \in M} p(m) \ln p(m) + \sum_{m \in M} p^e(m \mid \boldsymbol{l}^e, S^e) \ln p^e(m \mid \boldsymbol{l}^e, S^e), \tag{7}$$

where S^e is the set of qualitative states observed in e, and $\boldsymbol{l}^e$ the vector of measurements of landmark values. Notice that no temporal ordering of the qualitative states S^e is specified, unlike for the qualitative behavior b^e occurring in (5). By disregarding to a greater or lesser extent the information provided by the semi-quantitative behavior of the system, the criteria (6) and (7) may underestimate the discriminatory potential of an experiment (Section 6.2).

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

$$\Delta J(e) = E\big[\Delta H(\boldsymbol{l}^e, b^e)\big]. \tag{8}$$

The expected value of ΔH is computed by averaging over the possible experimental outcomes as predicted by the competing models. This is not easy to achieve in the case of the model discrimination problems we are interested in. First, the models in M are semi-quantitative differential equations and each SQDE may predict several semi-quantitative behaviors for the same experiment. Second, the measurements taken in the experiments are noisy, and therefore best described as random variables with a certain probability distribution.

Existing work in statistics and MBD does not provide much help in actually computing ΔJ in its general form (8). Statistical criteria are easily computable, but they are restricted to (algebraic) equations with exact numerical values

for the parameters and initial conditions. MBD approaches provide some help in calculating ΔJ from qualitative models in the case of measurement selection [20]. But they do not provide generic procedures to calculate ΔJ from the information contained in semi-quantitative differential equation models, a problem that is far from straightforward in general.

In Section 3.2, we derive an expression for the expected information increment applying to the discrimination of competing semi-quantitative models of dynamical systems. This expression involves probability distributions that can be estimated from semi-quantitative simulations of the experimental system. In Sections 3.3 and 3.4, we propose a procedure for the actual computation of these probability distributions.

3.2. Expected information increment of experiment

Recall that the information increment ΔH of an experiment e is defined by (5), while the expected information increment is given by (8). For clarity of presentation, we assume for the moment that only a single landmark l of observed variable y is measured in experiment e. The measured value is denoted by l^e. Obviously, the landmark value is included in the domain of y, i.e., $l^e \in D_y$ (or $l^e \in D$ for short).

The expected value of the information increment of an experiment e, $\Delta J(e)$, can then be computed in two stages: first, by averaging over the predicted qualitative behaviors, and second, by taking for each behavior the average of the model predictions for the landmark value. That is,

$$\Delta J(e) = E\big[\Delta H(l^e, b^e)\big] = \sum_{b \in B^e} E\big[\Delta H(l^e, b)\big] p^e(b), \tag{9}$$

where B^e is the set of possible qualitative behaviors b of the system in e, as predicted by the competing models, and $p^e(b)$ the probability that the system exhibits behavior b in e. $E[\Delta H(l^e, b)]$ is the expected value of ΔH relative to the conditional distribution of the landmark value, given b is the system behavior. That is,

$$E\big[\Delta H(l^e, b)\big] = \int_{l \in D} \Delta H(l, b) f^e(l|b)\, \mathrm{d}l,$$

where $f^e(l|b)$ is the conditional probability density function (pdf) of the landmark value, given that b is the behavior of the system in experiment e. Substituting the expression for $E[\Delta H(l^e, b)]$ in (9), we obtain for the expected information increment of e:

$$\Delta J(e) = \sum_{b \in B^e} \int_{l \in D} \Delta H(l, b) f^e(l|b) p^e(b)\, \mathrm{d}l. \tag{10}$$

According to the law of total probability, we can express the probability of b as a weighted sum of the conditional behavior probabilities $p^e(b|m)$:

$$p^e(b) = \sum_{m \in M} p^e(b|m) p(m), \tag{11}$$

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

$$f^e(l|b) = \sum_{m \in M} f^e(l|b, m) p^e(b, m),$$

where $f^e(l|b, m)$ is the model-specific pdf of the landmark value given b is the system behavior in e, and m the correct model of the system. $p^e(b, m)$ is the probability that model m is correct, and b is the system behavior in e. According to Bayes' rule

$$p^e(b, m) = p^e(b|m) p(m).$$

Hence, $f^e(l|b)$ becomes:

$$f^e(l|b) = \sum_{m \in M} f^e(l|b, m) p^e(b|m) p(m). \tag{12}$$

The model-specific pdfs $f^e(l|b,m)$, and the conditional behavior probabilities $p^e(b|m)$, satisfy the following normalization conditions:

$$\int_{l\in D} f^e(l|b,m)\,\mathrm{d}l = 1, \tag{13}$$

$$\sum_{b\in B_m^e} p^e(b|m) = 1, \tag{14}$$

where B_m^e is the set of qualitative behaviors of the system in e, as predicted by m. Sections 3.3 and 3.4 consider the determination of the model-specific pdfs and behavior probabilities in detail.

In the following proposition, the probability density function $f^e(l|b)$, the probability $p^e(b)$, and the normalization conditions (3), (13), and (14) are used to derive an expression for the expected information increment ΔJ.

Proposition 1. *The expected information increment of an experiment $e \in E$ is given by*

$$\Delta J(e) = \sum_{m\in M} p(m) \sum_{b\in B_m^e} p^e(b|m) \int_{l\in D} f^e(l|b,m) \ln \frac{f^e(l|b,m)p^e(b|m)}{f^e(l|b)p^e(b)}\,\mathrm{d}l. \tag{15}$$

Proof. By substituting the definition of ΔH into (10), we get

$$\Delta J(e) = \sum_{b\in B^e} \int_{l\in D} \left\{ \sum_{m\in M} p^e(m|l,b) \ln p^e(m|l,b) - \sum_{m\in M} p(m)\ln p(m) \right\} f^e(l|b)p^e(b)\,\mathrm{d}l, \tag{16}$$

where

$$p^e(m|l,b) = \frac{f^e(l,b|m)p(m)}{f^e(l,b)} = \frac{f^e(l|b,m)p^e(b|m)p(m)}{f^e(l|b)p^e(b)}, \tag{17}$$

via Bayes' rule. Combining (16) and (17), we obtain:

$$\begin{aligned}
\Delta J(e) &= \sum_{b\in B^e} \int_{l\in D} \sum_{m\in M} p(m) \left\{ \frac{f^e(l|b,m)p^e(b|m)}{f^e(l|b)p^e(b)} \ln \frac{f^e(l|b,m)p^e(b|m)p(m)}{f^e(l|b)p^e(b)} - \ln p(m) \right\} f^e(l|b)p^e(b)\,\mathrm{d}l \\
&= \sum_{m\in M} p(m) \sum_{b\in B_m^e} p^e(b|m) \int_{l\in D} f^e(l|b,m) \ln \frac{f^e(l|b,m)p^e(b|m)}{f^e(l|b)p^e(b)}\,\mathrm{d}l \\
&\quad + \sum_{m\in M} p(m)\ln p(m) \sum_{b\in B_m^e} p^e(b|m) \int_{l\in D} f^e(l|b,m)\,\mathrm{d}l - \sum_{m\in M} p(m)\ln p(m) \sum_{b\in B_m^e} \int_{l\in D} f^e(l|b)p^e(b)\,\mathrm{d}l.
\end{aligned}$$

Using the normalizations (3), (13), and (14) we obtain

$$\sum_{m\in M} p(m)\ln p(m) \sum_{b\in B_m^e} p^e(b|m) \int_{l\in D} f^e(l|b,m)\,\mathrm{d}l = \sum_{m\in M} p(m)\ln p(m),$$

$$\sum_{m\in M} p(m)\ln p(m) \sum_{b\in B_m^e} \int_{l\in D} f^e(l|b)p^e(b)\,\mathrm{d}l = \sum_{m\in M} p(m)\ln p(m).$$

With these equalities, 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 information increment. The *optimal* discriminatory experiment is the experiment 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 Proposition 1 provides an expression for ΔJ when all models predict the same qualitative behavior [61].

Corollary 1. *Let all models in M predict the same qualitative behavior b of the system in some experiment $e \in E$. Then* (15) *takes the form*

$$\Delta J(e) = \sum_{m \in M} p(m) \int_{l \in D} f^e(l|b,m) \ln \frac{f^e(l|b,m)}{f^e(l|b)} \, \mathrm{d}l. \tag{18}$$

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

$$\Delta J(e) = \sum_{m \in M} p(m) p^e(b|m) \int_{l \in D} f^e(l|b,m) \ln \frac{f^e(l|b,m) p^e(b|m)}{f^e(l|b) p^e(b)} \, \mathrm{d}l.$$

Moreover, since each model predicts a single qualitative behavior b, the conditional probability of this behavior equals 1 for all models. That is, $p^e(b|m) = 1$ for all $m \in M$. Consequently, the total probability of the behavior is also 1:

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

After replacing this value in the above expression for $\Delta J(e)$, the result (18) is obtained. □

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

Corollary 2. *Assume that, for a given experiment $e \in E$, each model $m \in M$ predicts a different set of qualitative behaviors. Then,*

$$\Delta J(e) = -\sum_{m \in M} p(m) \ln p(m). \tag{19}$$

Proof. Let $m' \in M$ and $m \neq m'$. From the given assumptions we have $p^e(b|m') = 0$, if $b \in B^e_m$. This fact, together with the expression for $f^e(l|b)$ given in (12), yields

$$f^e(l|b) p^e(b) = \sum_{m' \in M} f^e(l|b,m') p^e(b|m') p(m') = f^e(l|b,m) p^e(b|m) p(m),$$

for all $b \in B^e_m$. Taking this result into account, (15) takes the form

$$\Delta J(e) = \sum_{m \in M} p(m) \sum_{b \in B^e_m} p^e(b|m) \int_{l \in D} f^e(l|b,m) \ln \frac{1}{p(m)} \, \mathrm{d}l,$$

or equivalently,

$$\Delta J(e) = -\sum_{m \in M} p(m) \ln p(m) \sum_{b \in B^e_m} p^e(b|m) \int_{l \in D} f^e(l|b,m) \, \mathrm{d}l.$$

By applying the normalization conditions (13) and (14), expression (19) is obtained. □

Note that (19) 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) also applies when the landmark to be measured is a function of the landmarks of the observed variables (Section 2.2). In addition, it is easily generalizable to the case that multiple landmarks are measured. In that case we have to substitute the probability distributions by joint probability distributions, and the integral by a multiple integral for all landmarks. In [21] it is described how the entropy criterion can be generalized so as to take into account the varying costs of experiments.

3.3. Estimation of probability density functions

The computation of $\Delta J(e)$ requires the model-specific probability density functions $f^e(l|b,m)$ of the landmark values. The determination of $f^e(l|b,m)$ must be based on the predictions of m. Moreover, it is important that the definition $f^e(l|b,m)$ also incorporates the measurement uncertainty of the landmark.

Let l be a landmark of an observed variable y measured in experiment e, typically a maximum, minimum, or equilibrium value of the variable. To account for experimental noise, the measurement of l is assumed to be a random variable with mean l^e and a probability density function $h^e(l|l^e, b^e)$ of known analytic form (uniform, normal, ...) over the interval $D \subseteq \mathbb{R}$ [28].

The functions $f^e(l|b,m)$ required for the computation of $\Delta J(e)$ are the model-specific estimations of the pdf $h^e(l|l^e, b^e)$. In the case of quantitative models, the single point-value prediction l^e_m of the landmark value provided by m is usually interpreted as the mean of the measurement [28]. Hence, $f^e(l|b,m)$ can be defined as equal to $h^e(l|l^e_m, b^e_m)$. In the case of semi-quantitative differential equations, however, the predictions of landmark values are intervals. Consequently, a different approach for estimating the functions $f^e(l|b,m)$ is needed.

We assume that the predicted interval refers to the mean of the landmark value and that the probability density function $g^e(x|b,m)$ of the mean within the predicted interval is known. $g^e(x|b,m)$ specifies how the mean of the landmark measurement is distributed, given that b is the system behavior in e and m the correct model of the system. Then the pdf $f^e(l|b,m)$ of the landmark can be defined as the expected value of h^e:

$$f^e(l|b,m) = \int_{x \in D} h^e(l|x,b), g^e(x|b,m)\,\mathrm{d}x. \tag{20}$$

In other words, we define $f^e(l|b,m)$ to be the average of the possible values $h^e(l|x,b)$ weighted by their corresponding probabilities $g^e(x|b,m)$.

The model-specific pdfs of the mean of the landmark $g^e(x|b,m)$ can be obtained directly from the model predictions. Recall that the predictions of the competing models for the landmark values, obtained by QSIM, Q2, and Q3, have the form of intervals (Section 2.3). Since nothing is known about the distribution within a predicted interval, we assume a uniform distribution. More specifically, if $L^e_m \subseteq D$ is the interval predicted by m for landmark l, then $g^e(x|b,m) = 1/|L^e_m|$ for $x \in L^e_m$, and 0 otherwise, where $|\cdot|$ denotes interval length.

Assume for the moment that the measured value of the landmark is uniformly distributed on the confidence interval of the measurement. That is, $h^e(l|x,b) = 1/\varepsilon$, for $l \in [x - \varepsilon/2, x + \varepsilon/2]$, with ε the size of the confidence interval. The integral in (20) can then be computed exactly. More specifically, we have

$$f^e(l|b,m) = \begin{cases} \frac{l - \underline{L^e_m} + \varepsilon/2}{\varepsilon |L^e_m|}, & l \in [\underline{L^e_m} - \varepsilon/2, \underline{L^e_m} + \varepsilon/2], \\ \frac{1}{|L^e_m|}, & l \in [\underline{L^e_m} + \varepsilon/2, \overline{L^e_m} - \varepsilon/2], \\ \frac{-l + \overline{L^e_m} + \varepsilon/2}{\varepsilon |L^e_m|}, & l \in [\overline{L^e_m} - \varepsilon/2, \overline{L^e_m} + \varepsilon/2], \\ 0, & l \notin [\underline{L^e_m} - \varepsilon/2, \overline{L^e_m} + \varepsilon/2], \end{cases} \tag{21}$$

where $\underline{L^e_m}$ and $\overline{L^e_m}$ are the lower and the upper bound of the interval prediction L^e_m, respectively. Fig. 6 shows the estimate of the model-specific pdf of x^*, the equilibrium value of the amount of biomass in behavior b_1 as predicted by the Monod model (Fig. 3).

The expression for $f^e(l|b,m)$ defined above assumes that the mean of the landmark value is uniformly distributed in the predicted interval. Better estimations of the probability density functions $g^e(l|b,m)$, and hence of $f^e(l|b,m)$, can be obtained in the following way. Let $\boldsymbol{\theta}$ be the vector of landmarks corresponding to the parameters and initial conditions of m, given at the beginning of the experiment. Because of imprecise and incomplete knowledge about the system, the landmark values $\boldsymbol{\theta}$ in e are approximated by a corresponding interval vector $\boldsymbol{\Theta}^e$. We partition $\boldsymbol{\Theta}^e$ into vectors of intervals $\boldsymbol{\Theta}^e_j \subseteq \boldsymbol{\Theta}^e$, for each of which we specify the joint probability $p(\boldsymbol{\theta} \in \boldsymbol{\Theta}^e_j)$. This results in an approximation of the probability distribution of $\boldsymbol{\theta}$ in $\boldsymbol{\Theta}^e$. Obviously, if we do not know anything about the probability

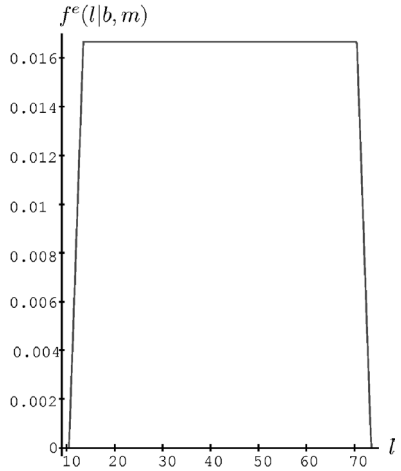


Fig. 6. Plot of the function $f^e(l|b,m)$ for the equilibrium amount of biomass (labeled x^*) of the behavior b_1 predicted by the Monod model (Fig. 3). The predicted interval for the equilibrium x^* is [11.98, 71.99]. (The size of the confidence interval ε has been set equal to 3.)

distributions of the parameters and initial conditions, then we assume that they are independent and uniform on the corresponding intervals. That is,

$$p\big(\boldsymbol{\theta} \in \boldsymbol{\Theta}^e_j\big) = \prod_{i \in I} p\big(\theta_i \in \Theta^e_{ji}\big),$$

where I is the set of indices of the landmarks corresponding to the parameters and initial conditions, and

$$p\big(\theta_i \in \Theta^e_{ji}\big) = \big|\Theta^e_{ji}\big| / \big|\Theta^e_i\big|,$$

with Θ^e_i denoting the interval bounds on θ_i in e.

We now simulate m for each of the vectors $\boldsymbol{\Theta}^e_j$. As a result, we obtain a set of (usually overlapping) intervals $L^e_{mj} \subseteq L^e_m$ bounding the landmark values. For each of the predicted intervals L^e_{mj} we again assume uniform distribution of the mean. That is, $g^e_j(x|b,m) = 1/|L^e_{mj}|$ for $x \in L^e_{mj}$, and 0 otherwise. This results in the following estimate of the model-specific pdf $g^e(x|b,m)$:

$$g^e(x|b,m) = \sum_{j \in J} p\big(\boldsymbol{\theta} \in \boldsymbol{\Theta}^e_j\big) g^e_j(x|b,m),$$

where J is the set of indices of the subintervals of $\boldsymbol{\Theta}^e$. By subdividing $\boldsymbol{\Theta}^e$ into ever finer subintervals $\boldsymbol{\Theta}^e_j$, we obtain progressively better estimates of $g^e(x|b,m)$, and hence of $f^e(l|b,m)$.

Consider, for instance, the Monod model given in Fig. 2. Assume the landmarks corresponding to the parameters Y, s_{in}, m, and k_s are distributed in the corresponding intervals (3) 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 $g^e(x|b,m)$ of the equilibrium value x^* of the biomass. In comparison with Fig. 6, where a uniform distribution of the mean in the predicted interval for x^* was assumed, a refined estimate of $f^e(l|b,m)$ is obtained (Fig. 7(b)).

Because the above approach arrives at more precise estimates of the model-specific pdfs, the prediction of optimal discriminatory experiments may be improved as well. But the improved predictions come at higher computational costs. In the example above, for instance, 1440 simulation runs were necessary to obtain the more precise estimate of (20). For more complex models involving many parameters, the benefit of the increased precision may not outweigh the computational costs (Section 6.3).

The above approach can be straightforwardly generalized to the case that multiple landmarks are taken into account. For computational simplicity, we assume that they are independent. Hence, the corresponding joint pdf is obtained by the product of the pdfs of the individual measured landmarks.

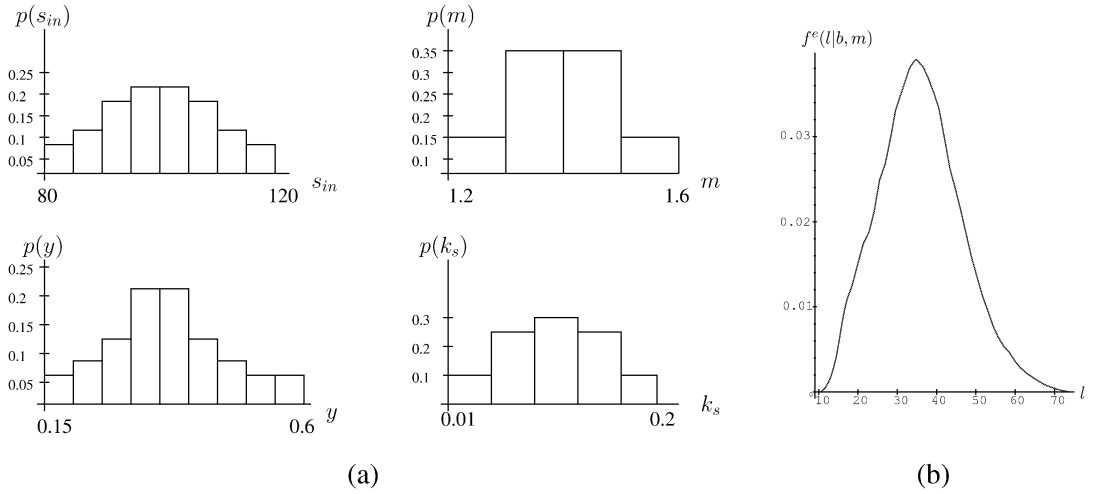


Fig. 7. (a) Approximate probability distributions of the landmarks corresponding to the parameters y, s_{in}, m, and k_s of the Monod model, given in the form of histograms. (b) Estimation of the probability density of x^* (behavior b_1 of the Monod model in Fig. 3) obtained by taking into account the distributions in (a).

3.4. Estimation of behavior probabilities

To compute $\Delta J(e)$, the model-specific behavior probabilities $p^e(b|m)$ must also be known. Obviously, if b is not predicted by m for experiment e, that is, $b \notin B_m^e$, then $p^e(b|m) = 0$. If b is one of the behaviors predicted by m, that is, $b \in B_m^e$, an estimate of $p^e(b|m)$ can be obtained by means of the following approach.

As before, let $\boldsymbol{\theta}$ be the vector of landmarks corresponding to the parameters and initial conditions of model m, given at the beginning of the experiment, and let $\boldsymbol{\Theta}^e$ be the vector of interval bounds on $\boldsymbol{\theta}$ in e. Furthermore, let $\boldsymbol{\Theta}_b^e \subseteq \boldsymbol{\Theta}^e$ be the vector of interval bounds on $\boldsymbol{\theta}$ giving rise to behavior $b \in B_m^e$, as obtained from the interval constraint propagation algorithms of Q2 and Q3. The latter programs are launched during simulation to refine, for each of the behaviors, the intervals $\boldsymbol{\Theta}^e$ specified in the model.

Assume the probability distributions of the parameters and initial conditions are given. Having determined the interval vector $\boldsymbol{\Theta}_b^e$, for each $b \in B_m^e$, we define:

$$r^e(b|m) = p\big(\boldsymbol{\theta} \in \boldsymbol{\Theta}_b^e\big),$$

where $p(\boldsymbol{\theta} \in \boldsymbol{\Theta}_b^e)$ is the probability that $\boldsymbol{\theta}$ lies in $\boldsymbol{\Theta}_b^e$. If we assume that the parameters and initial conditions are independently distributed, then

$$r^e(b|m) = \prod_{i \in I} p\big(\theta_i \in \Theta_{bi}^e\big),$$

where I is the set of indices of the landmarks corresponding to parameters and initial conditions, and $p(\theta_i \in \Theta_{bi}^e)$ the probability that the value of θ_i lies in the interval Θ_{bi}^e. The model-specific probability of behavior b is now estimated by normalizing the $r^e(b|m)$s:

$$p^e(b|m) = \frac{r^e(b|m)}{\sum_{b' \in B_m^e} r^e(b'|m)}. \tag{22}$$

If the probability distributions of the θ_is are unknown, we assume that they are uniform. That is,

$$p\big(\theta_i \in \Theta_{bi}^e\big) = \frac{|\Theta_{bi}^e|}{|\Theta_i^e|}.$$

Intuitively, $p^e(b|m)$ is thus given by the fraction of the parameter volume giving rise to b. Notice that refined estimates for the behavior probabilities can be developed in a similar way as in Section 3.3. But this will be omitted here.

Consider the behaviors b_1 and b_2 in Fig. 3, generated by simulation of the Droop model. b_1 is obtained for the initial cellular quota $q_0 \in [1.92, 7.62]$, 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 $r(b_1|\mathrm{D}) = 0.9468$ and $r(b_2|\mathrm{D}) = 0.1761$, where D stands for the Droop model. After normalization, we obtain the probability estimates $p(b_1|\mathrm{D}) = 0.8432$ and $p(b_2|\mathrm{D}) = 0.1568$.

Alternative approaches for estimating the probabilities of qualitative behaviors have been proposed by Berleant et al. [3] and Leitch and Shen [44]. In [3] the interval bounds on the parameters and the initial conditions are used to derive a lower and an upper bound on the probability of a qualitative behavior. We have not adopted this approach here, because the computation of the expected information increment requires that the model-specific behavior probabilities are real values. Leitch and Shen [44] 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 the 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 is more likely to occur than b'), but does not derive a quantitative estimate of their relative priorities (i.e., by how much b is more likely to occur than b').

3.5. Properties of the expected information increment

The criterion ΔJ has some important properties which are summarized in the theorems below.[3]

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 measurements made in an experiment will help us decide between the models.

Theorem 1. *The expected information increment of any experiment $e \in E$ is nonnegative*:

$$\Delta J(e) \geqslant 0. \tag{23}$$

Equality holds if all models have the same predictions.

Assume that in an experiment e the landmark values l_1 and l_2 are determined. Theorem 2 below shows that ΔJ has another desirable property: measuring both l_1 and l_2 while performing e is expected to be more informative than measuring only one of the landmarks.

Theorem 2. *Let $\Delta J(e)_1 = E[\Delta H(l_1^e, b^e)]$ be the expected information increment of $e \in E$ when landmark l_1 is measured, and $\Delta J(e)_{1+2} = E[\Delta H(l_1^e, l_2^e, b^e)]$ the expected information increment of e when both landmarks l_1 and l_2 are taken into account. Then,*

$$\Delta J(e)_{1+2} \geqslant \Delta J(e)_1.$$

In summary, Theorem 2 says that if we measure two landmarks in an experiment, the expected information increment of the experiment will be higher than if we measure only one. Or equivalently, we expect to better discriminate between the models when we observe more.

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.

[3] The proofs of the theorems are included in Appendix A.

4. Algorithm for model discrimination

4.1. 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 $m \in M$ is correct, if its probability estimate is above a predefined threshold value α, $0 < \alpha \leqslant 1$.

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

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

Step 1 Determine the predicted behaviors of the system for every $m \in M$ 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, $\boldsymbol{l}^e$.
Step 4 Compute the *a posteriori* probabilities $p^e(m|\boldsymbol{l}^e, b^e)$ of the models. Set $p(m)$ to $p^e(m|\boldsymbol{l}^e, b^e)$.
Step 5 Refine the interval bounds on the parameters of the models in M.
Step 6 Remove e from E.

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

In the first step of the algorithm, the competing models $m \in M$ are simulated by QSIM, Q2, and Q3 (Section 2.3) 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 $\Delta J(e)$ is computed for every experiment $e \in E$. The computation of $\Delta J(e)$ exploits the predictions of the system behavior in e obtained from the models in M. More specifically, the predictions are used to determine estimations of the model-specific probability densities and behavior probabilities (Sections 3.3 and 3.4). The experiment yielding the highest value of the expected information increment is performed in step 3. The results of the experiment, namely the qualitative behavior b^e of the observed variables and the measurements of the landmarks $\boldsymbol{l}^e$, are used in step 4 to compute the posterior model probabilities via Bayes' rule:

$$p^e(m|\boldsymbol{l}^e, b^e) = \frac{f^e(\boldsymbol{l}^e|b^e, m)\, p^e(b^e|m)\, p(m)}{f^e(\boldsymbol{l}^e|b^e)\, p^e(b^e)}.$$

$f^e(\boldsymbol{l}^e|b^e, m)$ is the model-specific pdf (Section 3.3) evaluated at the measurements of the landmarks $\boldsymbol{l}^e$, and $p^e(b^e|m)$ is the probability of behavior b^e conditional to model m (Section 3.4). The total behavior probability $p^e(b^e)$ and the value of the conditional pdf $f^e(\boldsymbol{l}^e|b^e)$ are computed as in (11) and (12), respectively.

The measurements of the landmarks $\boldsymbol{l}^e$ are also used, in step 5, to refine the interval bounds on the parameters of the models in M, using the interval constraint propagation algorithms of Q2 and SQCA. In the last step of the algorithm, the performed experiment is removed from the set of possible experiments E.

4.2. 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, on top of existing implementations of QSIM, Q2, and Q3 [27, 58], and contains approximately 1200 lines of code.[4] 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 density 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

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

specify the values of the probability threshold α and the type of distribution of the measurements. The implementation has been used in the application described in Section 5.

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 studied 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 *versus* when experiments are selected at random. For this purpose, the following strategy has been adopted [58,61]. First, one of the competing models has been arbitrarily selected. Exact parameter values have been randomly chosen from the corresponding interval ranges. The resulting quantitative model has been used to make exact predictions of the system behavior by numerical simulation.[5] The predicted landmark values have been treated as the mean of the measurements. Finally, using the above 'measurements', the algorithm for model discrimination has been applied a number of times for different values of ε and α.

For this performance study, we have used six competing SQDE models of a mass-spring system, making different assumptions on the spring and friction forces. The results, reported in [58,61], have shown that, as expected, the average number of experiments performed is higher in the case of random experiment selection. 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 landmarks are taken into account. This confirms the intuition that when more measurements are made in an experiment, that is, when more information about the system behavior becomes available, better discrimination is achieved (Theorem 2). Interestingly, we also found, at least for the mass-spring example, that the use of the refined estimates of the pdfs described in Section 3.3 did not lead to a significant reduction of the number of experiments needed for model discrimination.

5. Application: discrimination of models of phytoplankton growth

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

5.1. Biological background

Phytoplankton are microscopic plants playing a key role in marine ecosystems. Many phytoplankton species exist, each with a characteristic size, shape, and growth properties [33]. Like terrestrial plants, phytoplankton contain 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 require nutrients for their growth. 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.

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 the *chemostat* (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 unicellular 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 a new equilibrium is observed. Once the system has reached this equilibrium, a further dilution rate experiment can be performed by changing the value of D again and observing the system evolve towards

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

the next 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 behavior of the system. However, the data thus obtained are noisy, as in the case of most biological systems [7]. For instance, phytoplankton biomass is often estimated by phytoplankton biovolume, which is difficult to measure at high precision. Furthermore, the measurements of the remaining substrate concentration become unreliable for small concentration values, which are under the detection limit of the apparatus.

5.2. Modeling phytoplankton growth

A variety of models have been proposed for describing 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 [36]. In this study, four 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 [11,23,48] presented in Section 2, we consider two other models proposed by Contois [14] and Caperon and Meyer [12] (Fig. 8). The Contois model is a variant 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.

Because of coarse-grained 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 bounds on the parameter values have been

$\dot{x} = \mu(s)x - Dx$	$\dot{x} = \mu(s)x - Dx$
$\dot{s} = D(s_{in} - s) - \rho(s)x$	$\dot{s} = D(s_{in} - s) - \rho(s)x$
$\mu(s) = \mu_{max}\frac{s}{s+k_s}$	$\mu(s,x) = \mu_{max}\frac{s}{s+k_x x}$
$\rho(s) = \frac{1}{Y}\mu(s)$	$\rho(s,x) = \frac{1}{Y}\mu(s,x)$
(Monod)	(Contois)
$\dot{x} = \mu(q)x - Dx$	$\dot{x} = \mu(q)x - Dx$
$\dot{q} = \rho(s) - \mu(q)q$	$\dot{q} = \rho(s) - \mu(q)q$
$\dot{s} = D(s_{in} - s) - \rho(s)x$	$\dot{s} = D(s_{in} - s) - \rho(s)x$
$\mu(q) = \bar{\mu}(1 - \frac{k_q}{q})$	$\mu(q) = \mu_{max}\frac{q-k_q}{q-k_q+k_0}$
$\rho(s) = \rho_{max}\frac{s}{s+k_s}$	$\rho(s) = \rho_{max}\frac{s}{s+k_s}$
(Droop)	(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]

Fig. 8. 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 bounds are given in the accompanying table. The value of D depends on the experiment under consideration.

estimated from data obtained in preliminary experiments [5]. All experiments have been carried out at the Laboratoire d'Océanographie at Villefranche-sur-Mer (France).

5.3. Experiment selection approach

In order to discriminate between the four competing models discussed in the previous section, we have considered chemostat experiments consisting of changes in the value of the dilution rate D. In response to these changes, the population attains a new equilibrium value (Section 5.1). We have considered 21 possible 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 are bounded by intervals. During the experiments, the biomass x and the concentration of the remaining nutrient s can be measured. We consider the following landmarks of the observed variables, if they are above the detection limit of the apparatus: 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^*).

To obtain the predictions 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 qualitative behaviors for each of the models. This is a consequence of the complexity of the models and the large intervals for the parameters 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 in [6].

In the next sections, the simulation results are used to order the experiments according to their expected information increment. The criterion is computed as described in Section 3. In particular, we have used the standard instead of the refined estimates for the probability distributions, in order to reduce simulation costs. The predicted optimal discriminatory experiment is compared with the experiment that has actually been carried out. Next, data from the latter experiment is taken into account to update the model probabilities and to refine the interval bounds on the parameters. The equilibrium data are also used as initial conditions to simulate the models again and predict the next optimal dilution rate experiment.

5.4. 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 $D = 0$. In this section we investigate whether more discriminating experimental results would be obtained by setting D to another value. For this purpose, we have simulated the four models with QSIM, Q2, and Q3 for each of the dilution rate experiments. The initial conditions being used (Table 1) 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. 9). For all other values of D, the four models

Table 1
Initial conditions and applied dilution rate for the three experiments considered in the text. The refined intervals for the parameters after applying the constraint propagation algorithms in Q2 and SQCA are presented as well

Experiment	Initial conditions	Applied D	New parameter intervals
1	$x_0 \in [0.088, 0.165]$	$D = 0$	M&C: $Y \in [0.55, 0.6]$ D&CM: $k_q \in [1.6, 1.96]$
	$s_0 \in [50, 55]$		$q_0 \in [5.4, 7.6]$
2	$x_0 \in [30.7, 36.2]$ $s_0 \in [0, 0.01]$[a] $q_0 \in [1.39, 1.83]$[b]	$D = 0.95$	D: $k_q \in [1.65, 1.9]$ CM: $k_q \in [1.6, 1.74]$
3	$x_0 \in [25.3, 28.3]$ $s_0 \in [0, 0.01]$[a] $q_0 \in [2.82, 4.74]$[b]	$D = 0.45$	–

[a] Under the detection limit.
[b] Computed from an auxiliary model based on mass conservation [8].

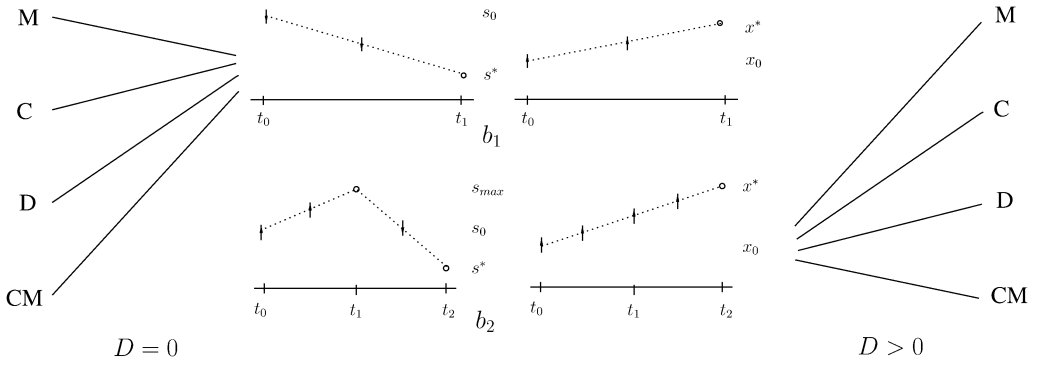


Fig. 9. Selection of initial experiment. Qualitative behaviors for the growth of *Dunaliella tertiolecta* predicted by the Monod (M), Contois (C), Droop (D), and Caperon–Meyer (CM) models for all dilution rate experiments, for the considered initial conditions. b_1 is predicted by the four models for the experiment $D = 0$, while b_2 is predicted for the other dilution rate experiments ($D > 0$).

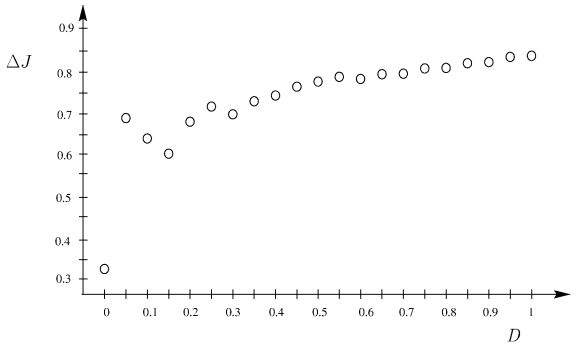


Fig. 10. Selection of initial experiment. Expected information increment ΔJ for each of the dilution rate experiments. $D = 1$ is the predicted optimal discriminatory experiment. Experiment $D = 0$ has been performed.

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. 9). 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 $p(b_2) = 1$.

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 x^*, s^* and s_{max}, as well as the behavior probabilities, the value of the expected information increment ΔJ has been computed for each of the dilution rate experiments (Fig. 10).

The figure shows that the optimal discriminatory experiment is $D = 1$, with expected information increment $\Delta J = 0.84$. As explained above, a single qualitative behavior b_2 is predicted. The interval predictions of the four models in this experiment are shown in Fig. 11(a). 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 experiment actually performed, $D = 0$, turns out to be the experiment with the lowest expected information increment ($\Delta J = 0.33$). Fig. 11(b) displays the predictions for x^* and s^* when $D = 0$. 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 results of experiment $D = 0$ to update the probability estimates of the models. As can be seen in Fig. 12(a), x and s asymptotically reach their equilibrium, in agreement with the predicted qualitative behavior b_1 and with the interval predictions of all models. As a result, none of the models can be ruled out. Using Bayes' rule, the *a posteriori* model probabilities can be computed:

$$p(\mathrm{M}) = 0.08, \qquad p(\mathrm{C}) = 0.08, \qquad p(\mathrm{D}) = 0.41, \qquad p(\mathrm{CM}) = 0.41. \tag{24}$$

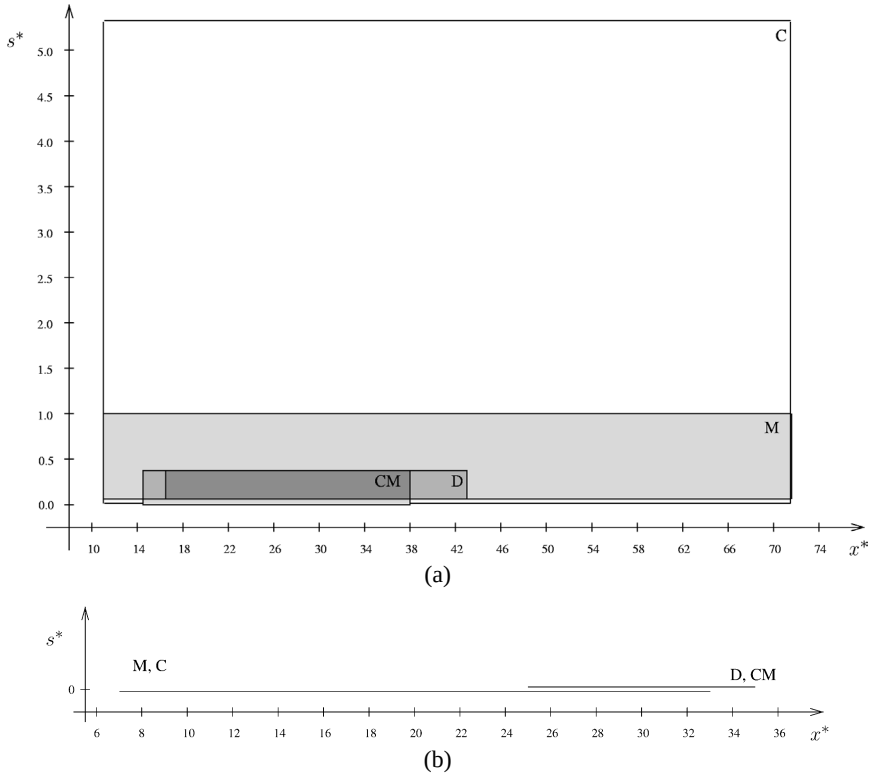


Fig. 11. Selection of initial experiment. Interval predictions of the four competing models for s^* and x^*. In (a), the predictions for the optimal discriminatory experiment $D = 1$ are given, while (b) shows the predictions for the experiment that has been actually carried out, $D = 0$. In the figures stronger shading designates more overlap between the predictions.

The *a posteriori* model probabilities give an estimation of both the quality of the fit between the model predictions and the observation, and the precision of these predictions. The predictions of the Droop and Caperon–Meyer models include the measurements for x^*. This is not the case for the predictions of the Monod and Contois models, which explains the lower probability of the latter.

The measurements allow the numerical bounds on the parameters in the competing models to be refined by the interval constraint propagation algorithms in Q2 and SQCA (Section 2.3), giving rise to the intervals in Table 1. These refined parameter intervals are used in the determination of the next optimal discriminatory experiment.

5.5. 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 $D = 0$ (Table 1).

Obviously, for $D = 0$ the state of the system does not change, because 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. 13). 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.

Using the model probabilities (24) and the simulation results, the expected information increment of each of the dilution rate experiments has been computed (Fig. 14). The results show that the experiment $D = 1$ is optimal with $\Delta J = 0.69$. Fig. 15 summarizes the model predictions for x^* and s^* (x_{min} and s_{max} have been omitted, because they are not much discriminatory). As can be seen, the predictions of the Monod and Contois models on one hand, and the Droop and Caperon–Meyer models on the other, do not overlap. Even when taking into account uncertainty in

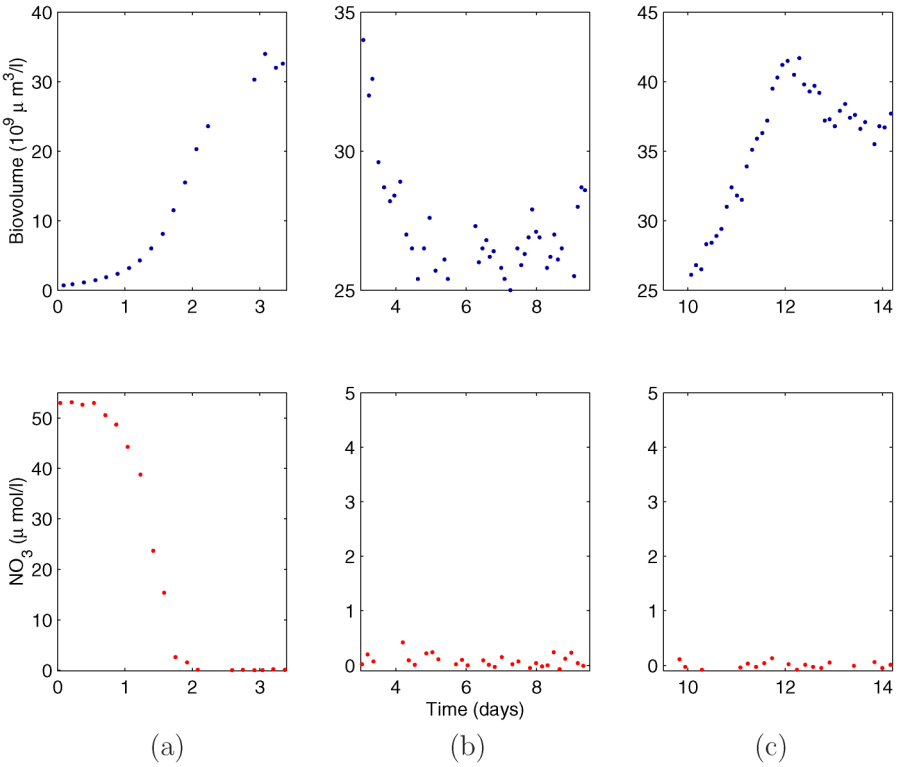


Fig. 12. Plot of the measurements of the temporal evolution of the biomass x (approximated by biovolume) and the substrate concentration s (NO_3) in (a) the initial experiment ($D = 0$), (b) the second experiment ($D = 0.95$), and (c) the third experiment ($D = 0.45$) (Bernard and Sciandra, unpublished data).

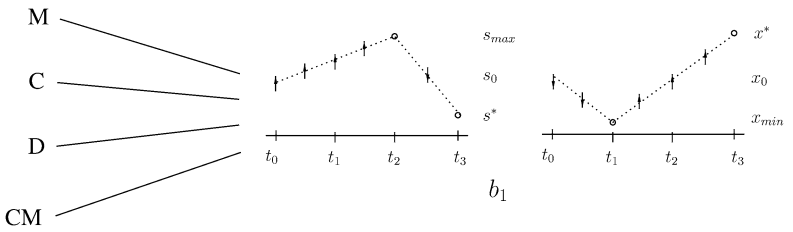


Fig. 13. Selection of second experiment. Qualitative behavior for the growth of *D. tertiolecta* predicted by the four competing models for every dilution rate experiment. For all D, $0 < D \leqslant 1$, a single qualitative behavior b_1 is predicted.

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. 12(b)). The curve agrees with the predicted qualitative behavior, while the equilibrium values (Table 1) provide the following *a posteriori* model probabilities:

$$p(\mathrm{M}) = 0, \qquad p(\mathrm{C}) = 0, \qquad p(\mathrm{D}) = 0.56, \qquad p(\mathrm{CM}) = 0.43. \tag{25}$$

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. As shown in Table 1, the experimental data refine the interval bounds for the parameter k_q.

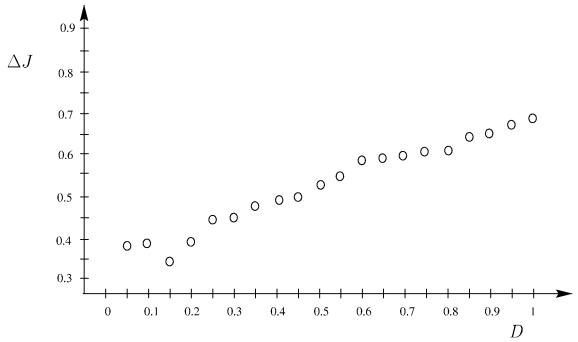


Fig. 14. Selection of second experiment. Expected information increment ΔJ for varying values of the dilution rate. Experiment $D = 1$ is the optimal discriminatory experiment, whereas experiment $D = 0.95$ has been performed.

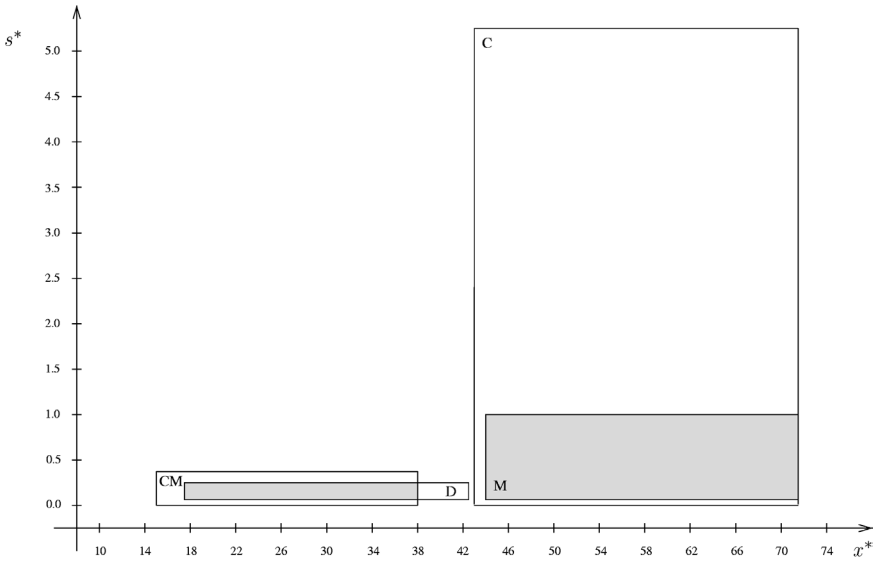


Fig. 15. Selection of second experiment. Interval predictions of the four competing models for the experiments for $D = 1$ (optimal discriminatory experiment). The predictions concern the landmarks x^* and s^*.

5.6. 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 the Caperon–Meyer models? As before, we simulate the models from the equilibrium attained in the previous step. For every experiment, each of the models predicts two possible qualitative behaviors for the biomass x (Fig. 16). For $D < 0.95$, the biomass is expected to increase asymptotically towards its equilibrium (behavior b_1), or to first pass through a maximum (behavior b_2). For $D > 0.95$, the models predict that x either decreases asymptotically to its new equilibrium (behavior b_3), or first passes through a minimum (behavior b_4). Estimation of the model-specific behavior probabilities, as explained in Section 3.3, 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) indicates that $D = 0.9$ is the optimal discriminatory experiment. In reality, $D = 0.45$ was performed. In that experiment x was found to reach its equilibrium after passing through a maximum (Fig. 12(c)). As a consequence, behavior b_1 is ruled out. The landmark values x_{max} and x^* were measured to lie within $[41.3, 43.7]$ 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:

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

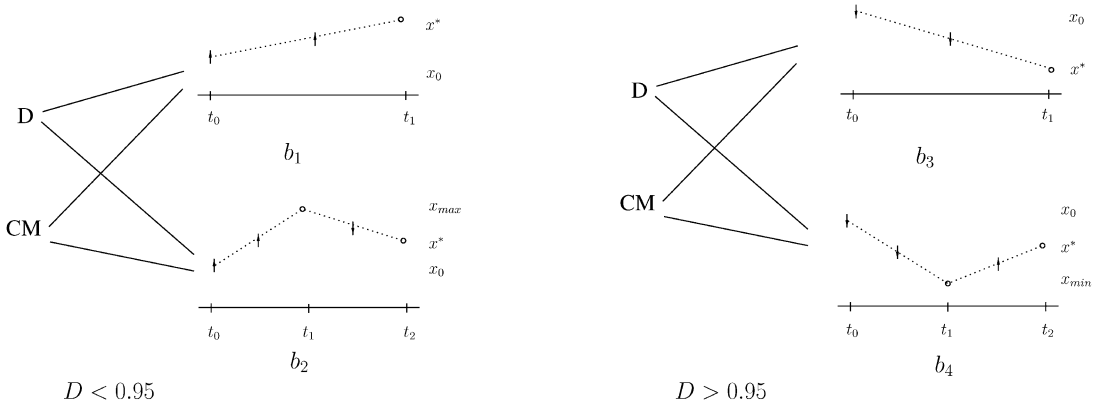


Fig. 16. Selection of third experiment. Qualitative behaviors for the growth of *Dunaliella tertiolecta* predicted by the four competing models for every dilution rate experiment. Behaviors b_1 and b_2 are predicted for the experiments $D < 0.95$, while behaviors b_3 and b_4 are predicted for the experiment $D > 0.95$.

That is, the experiment has not succeeded in discriminating between 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 are 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.

In summary, the application illustrates that the choice of experiments plays a crucial role in model discrimination. If we had chosen 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. This would have made the subsequent steps unnecessary. Also, if we had chosen an experiment with a low D value in the second experiment, the Monod and Contois models would probably not have been ruled out. This complements the results obtained in Section 4.2, where we showed by 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 actually performed instead of predicted to be optimally discriminatory. In the process, the interval bounds on one of the parameters in the remaining two models were considerably refined. Although the Monod model, and its variants like the Contois model, are widely used in practice, the results of our study demonstrate that they are not suitable to represent a broad range of phytoplankton growth conditions. Instead, the Droop model or the closely-related Caperon–Meyer model should be preferred. The difficulty to discriminate between the latter two models reflects their similarity. Further experiments, performed under conditions for which the predictions of the two models diverge, are needed to discriminate between the models.

6. Discussion and related work

The method described in this paper does not make any domain-specific assumptions, which guarantees that it can be used for the discrimination of 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 Figs. 4 and 5: the models, the experiments, the simulation techniques, the model discrimination criterion, and the model discrimination strategy. These assumptions are discussed below, in the context of related work.

6.1. Models

Most model discrimination methods have been developed for *quantitative* models, in particular algebraic equation and ordinary differential equation (ODE) models. These models allow precise numerical predictions of the system behavior to be made, but their use requires precise numerical values of the parameters and exact functional relations

between the variables. In many situations, the available information on the system is incomplete or imprecise, thus impeding the application of quantitative models. This is especially so in domains in which mathematical modeling has been less widespread thus far, like biology, medicine, and chemistry. The inability to quantitatively monitor the system behavior over time, another frequently-occurring problem, may further complicate the use of ODE models.

In response to these problems, researchers in qualitative reasoning (QR) have proposed qualitative models of dynamical systems, typically qualitative differential equation (QDE) models [17,26,56,62]. These models have been widely used in domains where ODE models are difficult to formulate. For model discrimination, the qualitative behaviors produced by QSIM and other qualitative simulators may not be sufficiently precise though. In addition, they experience upscaling problems due to the explosion of the number of qualitative behaviors, some of which may be spurious.

The *semi-quantitative* models used here strike a compromise between the precision of quantitative models and the applicability of qualitative models, in situations where the available information is incomplete or imprecise. More particularly, semi-quantitative differential equations (SQDEs) extend qualitative differential equations by specifying interval bounds on parameter values and numerical envelopes for monotonic functions [4,38]. The use of this weak numerical information also helps alleviate the upscaling problems experienced by qualitative simulation, in the sense that the number of qualitative behaviors generated may be reduced and some of the spurious behaviors ruled out. But upscaling problems still remain, partly due to the lack of suitable computer tools (Section 6.3).

6.2. Criteria for experiment selection

In order to discriminate among competing models of a dynamical system, we have to establish a criterion that allows ranking of the possible experiments according to their discriminatory potential. Criteria for the determination of optimal discriminatory experiments have been developed in statistics and in AI. They have been used for a wide range of problems in different domains of application, such as in biology and biotechnology (e.g., [15,37,55,57]), in physics (e.g., [16,45]), in chemical engineering (e.g., [1,25]), and in model-based diagnosis (e.g., [20,22,53]).

Most criteria for experiment selection are based on the principle of *maximum divergence* [2,25,28,29,35]. Intuitively speaking, this means that a measure for the difference in model predictions in response to different experimental conditions is defined, followed by a search for the conditions under which this measure is maximized. Here we have focused on one particular type of maximum divergence criterion, the *maximum entropy* criterion, but other examples can be found in the references given above. The basic idea underlying the maximum entropy criterion is to find the experiment with the highest *information increment* ΔH (4), that is, the experiment for which the difference in entropy before and after execution is maximal. The expected value ΔJ of this information increment, for each of the available experiments, can be determined from the model predictions.

The specific form that the maximum entropy criterion (4) takes, and therefore the resulting ranking of the experiments, is critically dependent on the characteristics of the behavior of the dynamical system we wish to consider. In this article, we have proposed a criterion, given by (5), that is specifically adapted to a semi-quantitative description of the dynamics: it takes into account the qualitative behavior of the system as well as interval bounds on the landmark values of the variables. This generalizes upon the maximum entropy criteria found in the statistical and MBD literature, given by (6) and (7), respectively. While the former omits the qualitative behavior of the system, the latter only considers qualitative states and not their temporal ordering in qualitative behaviors.

The use of the generalized maximum entropy criterion may increase the discriminatory potential of the method. Consider the case that two competing models predict the single observed variable of a system to reach an equilibrium value in an experiment. If the two models predict overlapping values for the equilibrium value, then the experiment will be assigned a low expected information increment according to (6). However, this is not appropriate when the models predict different qualitative behaviors of the system, e.g., an asymptotic increase *versus* decrease of the observed variable towards its equilibrium value. The criterion we propose here, given by (5), would in this case assign a high expected information increment for the experiment, as desired, by taking into account this predicted difference in qualitative behavior.

It can be argued that, in order to discriminate between two models, it is not necessary to take into account qualitative behaviors. In fact, information on the observed qualitative states only, leaving aside their temporal ordering in qualitative behaviors, would already suffice to refute one of the models. This idea is based on the insight that two models predicting different qualitative behaviors of a system will also differ on the possible qualitative states they

predict [54]. It underlies the maximum entropy criterion (7), proposed by Struss [53] in the context of MBD, generalizing upon earlier work by de Kleer and Williams [22] and de Kleer [20]. The criterion (7) has the advantage that it is computationally less expensive than criterion (5), because it does not require the transitions between qualitative states to be computed. However, it relies on the capability to observe the qualitative states predicted to be different for the two models. This may not be possible if only some of the variables are observed, a situation not uncommon in practice. In that case, an experiment might be assigned a low expected information increment according to (7), whereas it should receive a high expected information increment. The additional computational costs of criterion (5) may therefore pay off. They can increase the discriminatory potential by exploiting information on the time evolution of the system, in particular under conditions of limited measurability.

To be practically useful, the information increment (5), or rather its expected value (8), has to be computed from the predictions of the models. To this end, we have proposed a problem-independent computational framework appropriate for SQDE models, more specifically a framework based on semi-quantitative simulation. Several technical difficulties were solved to achieve this, notably the computation of the model-specific probability density functions $f^e(l|b, m)$ of the landmark values (Section 3.3) and of the model-specific behavior probabilities $p^e(b|m)$ (Section 3.4). We are not aware of any existing work addressing these problems. Methods in statistics deal with quantitative models only. While MBD methods allow qualitative or semi-quantitative models, they either do not specify computational procedures or develop procedures specific for a certain type of problem (see [20,22] for an example of the latter case).

The development of an expression for the expected information increment in Section 3.2 implicitly assumes that in determining the optimal next experiment we only need to look one step ahead. However, one can imagine that a better estimate of ΔJ would be obtained by evaluating the consequences of carrying out several consecutive experiments, in the limit proceeding until only a single model remains [22]. The increase in discriminatory potential of the optimal next experiment thus obtained, by using a *multi-step lookahead* instead of a *one-step lookahead* approach, needs to be carefully balanced against the additional computational costs involved. de Kleer et al. [21] show that, in the context of the diagnosis of electronic circuits, one-step lookahead leads to near-optimal results.

6.3. Simulation techniques

Simulation is a crucial step in the method for model discrimination. The expected information increment of each experiment is evaluated on the basis of the predictions derived from the competing models. As a consequence, the techniques should derive all possible predictions, and these predictions must be as precise as possible. Omitted predictions may cause a model to be falsely ruled out when the predictions derived from this model do not match the observation in an experiment. Precise predictions are important to efficiently discriminate between the models, since overestimation of the real solutions may, for instance, cause a model to be corroborated when it should be ruled out.

The techniques employed in this work satisfy the first requirement: QSIM, Q2, Q3, and SQCA have been proven sound (Section 2.3). As a consequence, models are only ruled out when they should be. But due to the incompleteness of the simulation techniques, spurious behaviors may be generated and the interval predictions may overestimate the real solutions. This leads to imprecise estimates of the model-specific probability density functions and behavior probabilities, and hence to imprecise estimates of ΔJ. Moreover, spurious predictions may lead to imprecise estimates of the posterior model probabilities.

The simulation techniques employed in this paper aim at obtaining interval predictions of the landmarks 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 the simulation. Within the same computational framework, Q2 also may refine interval bounds on parameter values by propagating the results of conducted experiments. The refinement of parameter values through the integration of measurements can be further improved by SQCA, which infers additional constraints from the comparison with models describing the system in previously performed experiments.

As noted previously, the increased precision of the predictions obtained by the above techniques has 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. These and similar problem make simulation the computational bottleneck of the approach presented in this article, especially in the case of large models and refined

estimates of the model-specific probability density functions (Section 3.3) and model-specific behavior probabilities (Section 3.4).

The upscaling problems are partly due to the lack of high-performance computer tools for semi-quantitative simulation. But even if efficient computer tools existed, there remains a fundamental trade-off between the precision of predictions on one hand and the computational costs on the other. In some cases the price of intensive simulation studies aimed at high-precision predictions may be too high, whereas in others the benefits of avoiding costly and time-consuming experiments may outweigh the computational costs.

6.4. Discriminatory experiments

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.

Sometimes it is difficult or even impossible to intervene with the system behavior. In such cases experiment selection for model discrimination is restricted to *measurement selection* [22,46,49]. In measurement selection, one determines the optimal measurement to perform next. Obviously, experiment selection goes beyond measurement selection in that one can determine not only the optimal measurement to perform but also the optimal experimental conditions to apply. In fact, measurement selection makes sense only when costs of individual measurements are taken into account. Indeed, as shown by Theorem 2 and the computer experiments mentioned in Section 4.2, if one measures more landmarks in an experiment, the expected information increment of the experiment increases. If measurements are costly, one can select the most informative subset that obeys specified cost limits, using techniques that are similar to those used for experiment selection.

Experiment selection starts with a set of available experiments. One can argue, however, that this provides no guarantee that the optimal experiment is included in this set. If ΔJ is known or suspected to be a non-smooth function, then the discretization of a continuous range of experimental conditions may introduce a bias, and *experiment design* may become an option. The basic idea of experiment design is to define the expected information increment ΔJ as a function of the experimental conditions that need to be fixed in the experiment. The problem of finding the optimal experiment is then mapped to an optimization problem: one has to find the experimental conditions for which the function reaches its maximum [1,13,25,28]. By determining optimal experimental conditions within a continuous range, experiment design offers a greater discriminatory capability. But this further increases the computational complexity of the model discrimination problem, since global optimization approaches require a large amount of simulation runs [34]. Given that simulation is the computational bottleneck in the discrimination of semi-quantitative models (Section 6.3), experiment design may become unfeasible in practice.

Even more ambitiously, experiment design can go beyond fixing a certain parameter value or initial condition, and apply to broader aspects of the experimental conditions and experimental set-up. A few explorations of this form of experiment design can be found in the AI literature, for instance in the work of Rajamoney [50] and Bradley et al. [10]. Generally speaking, experiment design thus conceived requires an explicit representation of the elements of the experimental set-up, as well as methods to construct a model of the experimental system from a domain theory and a description of the experimental set-up. This introduces issues in knowledge representation and automated reasoning that fall outside the scope of this article.

6.5. Model discrimination strategy

In order to discriminate a set of models, we have adopted a method that iterates the selection and execution of experiments, where the selection step is based on the expected information increment criterion (5). We have carried out an extensive simulation study to test the performance of this strategy in the case of the discrimination of semi-quantitative models of a simple mechanical system (Section 4.2). Under a wide range of conditions, this approach towards model discrimination turned out to be more efficient than the random choice of experiments, in the sense that the average number of experiments required to discriminate between the models is lower when criterion (5) is employed.

The iteration of the selection and execution of experiments amounts to a sequential strategy for model discrimination [22,30]. One could imagine another strategy, consisting first of the determination *a priori*, not of a single

discriminatory experiment, but of a sequence of experiments guaranteed to discriminate between all models. Once this experimental plan has been computed, it could be executed in a second step. As Struss [53] observes, the sequential approach can save computational time, because refuted models need not be simulated when determining further optimal experiments. In addition, it can render some experiments unnecessary when the models they are supposed to discriminate have already been ruled out by previous experiments.

Bradley and colleagues have proposed another strategy for a problem related to model discrimination: find the simplest model that is valid under a representative range of experimental conditions [10]. Instead of starting with an explicitly-defined set of possible models, the approach starts with the simplest model consistent with user-specified hypotheses and a domain theory. This model is tested against available observations using increasingly precise and sophisticated techniques to analyze the model, ranging from qualitative simulation to numerical parameter estimation, simulation, and bifurcation analysis. If the model fails to reproduce the data, a more detailed model is constructed. On the other hand, if the model succeeds in reproducing the data, further experiments to corroborate or refute the model are proposed and executed. The method terminates when a model is found that is globally valid across the range of experimental conditions. The idea of the hierarchical testing of model validity, postponing the computationally-intensive tests until the model has successfully passed simpler tests, might be profitably used for model discrimination as well. However, testing one model at a time may significantly raise computational costs if the model space is large, in particular when many simple models are inconsistent with the data. For model discrimination, the strategy adopted in this article, which consists in finding optimal experiments that refute as many models as possible at once, will in general be more efficient both computationally and experimentally.

7. Conclusions and further work

Different assumptions on the structure and behavior of a system may result in a number of competing models, all justifiable by the available observations. This gives rise to the problem of model discrimination, which occurs in many guises in practically every domain of science and engineering. To discriminate between the models new experiments are needed. Most experiment selection methods developed in statistics and in artificial intelligence are based on the information increment of an experiment, that is, the difference in entropy before and after the execution of the experiment. Existing model discrimination methods are limited in two respects. First, the information increment criteria being used are not well-adapted to the discrimination of models of dynamical systems under conditions of limited measurability. Second, there are no generic, domain-independent procedures for the computation of the expected information increment when the models are qualitative or semi-quantitative. Many applications, however, concern dynamical systems in which variables cannot be measured, unless with great difficulty. Moreover, the information required to specify numerical models is often absent.

The above situation has motivated the development of a method for the selection of experiments to discriminate between semi-quantitative models of dynamical systems. The method has been implemented on top of existing implementations of the qualitative and semi-quantitative simulation techniques QSIM, Q2, and Q3, which are used to derive model predictions. The method and its implementation are independent of a particular application, as they do not employ any domain-specific knowledge about the experimental system. 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.

The practical applicability of the method has been demonstrated on a real problem in the field of population biology. In particular, it has been shown that the method predicts the most informative experiments to discriminate between four 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 system behavior. Our results show that the widely-used Monod model and its variants are not valid for a broad range of phytoplankton growth conditions and that the Droop or Caperon–Meyer model should be preferred. In addition, the model discrimination process has led to refined interval bounds on a key parameter of these models.

Several directions for further work can be identified. First of all, the discussion in the previous section has brought to the fore the central role of semi-quantitative simulation in the model discrimination method. In particular, the use of simulation involves a fundamental trade-off between experimental and computational costs. On one hand, refinement of the interval bounds may lead the method to propose better discriminatory experiments, due to more precise

estimations of ΔJ. On the other hand, the application of more sophisticated simulation techniques may cause the time spent on computing ΔJ to explode. More efficient implementations of current semi-quantitative simulation techniques could relieve some of the upscalability problems. Another 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, while Monte Carlo simulation could be used to derive fast approximations of the probability distributions.

More generally, the method for model discrimination described in this paper can be integrated with other tools for computer-supported modeling, such as tools for model building, model validation, and model revision, thus giving rise to integrated environments for computer-supported modeling [10,19]. Model building, conceived of as either the composition of a model from a set of reusable model fragments or as the induction of a model from the observed behavior of the experimental system [40,51,63], may lead to a large set of competing models. The selection of optimal experiments will help in determining the correct model systematically and efficiently. The explicit representation of the domain knowledge and knowledge on the experimental set-up would also provide the necessary prerequisites for developing more ambitious forms of experiment design.

Acknowledgements

The authors would like to thank Jean-Luc Gouzé for his contributions to the work presented here and the reviewers for their comments on previous versions of the article. Most of the work described here 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 of the Programme d'actions intégrées franco-néerlandaises Van Gogh (dossier 99033).

Appendix A. Proofs of theorems

This appendix details the proofs of the theorems given in Section 3.5.

Theorem 5. *The expected information increment of any experiment $e \in E$ is nonnegative*:

$$\Delta J(e) \geqslant 0. \tag{A.1}$$

Equality holds if all models have the same predictions.

Proof. From Theorem 204 in [32] the following inequality can be established:

$$\sum_{b\in B_m^e} \int_{l\in D} \xi_m(l,b) f^e(l|b) p^e(b) \ln \xi_m(l,b)\, \mathrm{d}l$$
$$\geqslant \sum_{b\in B_m^e} \int_{l\in D} \xi_m(l,b) f^e(l|b) p^e(b)\, \mathrm{d}l \ln \frac{\sum_{b\in B_m^e} \int_{l\in D} \xi_m(l,b) f^e(l|b) p^e(b)\, \mathrm{d}l}{\sum_{b\in B_m^e} \int_{l\in D} f^e(l|b) p^e(b)\, \mathrm{d}l}, \tag{A.2}$$

where $p^e(b)$ is the probability of behavior $b \in B_m^e$, $f^e(l|b)$ is the conditional pdf of the landmark as given by (12), and $\xi_m(l,b)$ a bounded function. Equality holds only if $\xi_m(l,b)$ is constant, that is, if it is independent of a particular b and l. Define

$$\xi_m(l,b) = \frac{f^e(l|b,m) p^e(b|m)}{f^e(y|b) p^e(b)},$$

where $p^e(b|m)$ is the model-specific probability of $b \in B_m^e$, and $f^e(l|b,m)$ is the model-specific pdf of the landmark. Hence, we can rewrite the right-hand side of (A.2) as follows

$$\sum_{b\in B_m^e} \int_{l\in D} f^e(l|b,m) p^e(b|m)\, \mathrm{d}l \ln \frac{\sum_{b\in B_m^e} \int_{l\in D} f^e(l|b,m) p^e(b|m)\, \mathrm{d}l}{\sum_{b\in B_m^e} \int_{l\in D} f^e(l|b) p^e(b)\, \mathrm{d}l}. \tag{A.3}$$

Taking into account the normalization conditions

$$\int_{l\in D} f^e(l|b,m)\,\mathrm{d}l = 1 \quad \text{and} \quad \sum_{b\in B_m^e} p^e(b|m) = 1$$

given in (13) and (14), and using that $f^e(l|b) = \sum_{m'\in M} f^e(l|b,m')p^e(b|m')p(m')/p^e(b)$ from (12), we obtain

$$\sum_{b\in B_m^e}\int_{l\in D} f^e(l|b)p^e(b)\,\mathrm{d}l = \sum_{b\in B_m^e}\int_{l\in D}\sum_{m'\in M} f^e(l|b,m')p^e(b|m')p(m') = 1.$$

Consequently, the argument of the logarithmic function becomes 1, and the expression in (A.3) simplifies to zero. Hence,

$$\sum_{b\in B_m^e}\int_{l\in D} \xi_m(l,b)f^e(l|b)p^e(b)\ln\xi_m(l,b)\,\mathrm{d}l \geqslant 0.$$

Since the expression for the expected information can be rewritten in the form

$$\Delta J(e) = \sum_{m\in M} p(m)\sum_{b\in B_m^e}\int_{l\in D} \xi_m(l,b)f^e(l|b)p^e(b)\ln\xi_m(l,b)\,\mathrm{d}l,$$

we finally obtain

$$\Delta J(e) \geqslant 0.$$

Equality holds if $\xi_m(l,b)$ is constant for all m. This happens only if the $p^e(b|m)$s and the $f^e(l|b,m)$s are the same for all behaviors and models. That is, if all models give rise to identical predictions. □

Theorem 6. *Let $\Delta J(e)_1 = E[\Delta H(l_1^e, b^e)]$ be the expected information increment of some experiment $e \in E$ when landmark l_1 is measured, and $\Delta J(e)_{1+2} = E[\Delta H(l_1^e, l_2^e, b^e)]$ the expected information increment of e when both landmarks l_1 and l_2 are taken into account. Then,*

$$\Delta J(e)_{1+2} \geqslant \Delta J(e)_1.$$

Proof. In this theorem, we consider landmarks of a single behavior. The formula for the expected information increment therefore simplifies to the expression given in Corollary 1.

By Corollary 1, we have

$$\Delta J(e)_1 = \sum_{m\in M} p(m)\int_{l_1\in D_1} f^e(l_1|b,m)\ln\frac{f^e(l_1|b,m)}{f^e(l_1|b)}\,\mathrm{d}l_1. \tag{A.4}$$

Using the normalization condition

$$\int_{l_2\in D_2} f^e(l_2|l_1,b,m)\,\mathrm{d}l_2 = 1,$$

and the definition of the joint pdf of l_1 and l_2 as the product of the individual pdfs:

$$f^e(l_1,l_2|b,m) = f^e(l_2|l_1,b,m)f^e(l_1|b,m), \tag{A.5}$$

we can rewrite (A.4) as

$$\Delta J(e)_1 = \sum_{m\in M} p(m)\int_{l_1\in D_1}\int_{l_2\in D_2} f^e(l_1,l_2|b,m)\ln\frac{f^e(l_1,l_2|b,m)}{f^e(l_2|l_1,b,m)f^e(l_1|b)}\,\mathrm{d}l_1\,\mathrm{d}l_2. \tag{A.6}$$

Given that

$$f^e(l_1,l_2|b) = f^e(l_2|l_1,b)f^e(l_1|b), \tag{A.7}$$

Eq. (A.6) can be written as

$$\Delta J(e)_1 = \sum_{m\in M} p(m) \int_{l_1\in D_1} \int_{l_2\in D_2} f^e(l_1, l_2|b, m) \ln \frac{f^e(l_1, l_2|b, m)}{f^e(l_1, l_2|b)} \,\mathrm{d}l_1 \,\mathrm{d}l_2 - \sum_{m\in M} p(m) \int_{l_1\in D_1} \int_{l_2\in D_2} f^e(l_1, l_2|b, m) \ln \frac{f^e(l_2|l_1, b, m)}{f^e(l_2|l_1, b)} \,\mathrm{d}l_1 \,\mathrm{d}l_2. \tag{A.8}$$

Notice that the first expression on the right-hand side of (A.8) equals $\Delta J(e)_{1+2}$, by Corollary 1. The second expression can be rewritten by means of (A.5), using the function ζ defined as follows:

$$\zeta(l_1) = \sum_{m\in M} p(m) \int_{l_2\in D_2} f^e(l_2|l_1, b, m) \ln \frac{f^e(l_2|l_1, b, m)}{f^e(l_2|l_1, b)} \,\mathrm{d}l_2. \tag{A.9}$$

This results in

$$\Delta J(e)_1 = \Delta J(e)_{1+2} - \int_{l_1\in D_1} \zeta(l_1) f^e(l_1|b, m) \,\mathrm{d}l_1. \tag{A.10}$$

In fact, $\zeta(l_1)$ is an expected information increment of the form (18), so that $\zeta(l_1) \geqslant 0$ by Theorem 1. As a consequence, the integral in the second expression on the right-hand side of (A.10) is greater than or equal to zero, since it represents the expected value of ζ. This results in

$$\Delta J(e)_{1+2} \geqslant \Delta J(e)_1. \qquad \square$$

References

[1] S. Asprey, S. Macchietto, Designing robust optimal dynamic experiments, J. Process Control 12 (2002) 545–556.
[2] A. Atkinson, D. Cox, Planning experiments for discriminating between models, J. Roy. Statist. Soc. 36 (1974) 321–348.
[3] D. Berleant, A. Chandra, K. Bognaes, C.-G. Liaw, L. Sheng, 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, ACM Press, New York, 1992, pp. 883–889.
[4] D. Berleant, B. Kuipers, Qualitative and quantitative simulation: Bridging the gap, Artificial Intelligence 95 (2) (1997) 215–256.
[5] O. Bernard, Étude expérimentale et théorique de la croissance de *Dunaliella tertiolecta* soumise à une limitation variable de nitrate, Ph.D. thesis, Université Pierre & Marie-Curie, Paris, France, 1995.
[6] O. Bernard, J.-L. Gouzé, Global qualitative description of a class of nonlinear dynamical systems, Artificial Intelligence 136 (1) (2002) 29–59.
[7] O. Bernard, G. Malara, A. Sciandra, The effects of a controlled fluctuating nutrient environment on continuous cultures of phytoplankton monitored by computers, J. Experimental Marine Biology and Ecology 197 (2) (1996) 263–278.
[8] O. Bernard, G. Sallet, A. Sciandra, Nonlinear observers for a class of biological systems: Application to validation of a phytoplanktonic growth model, IEEE Trans. Automat. Control 43 (8) (1998) 1056–1065.
[9] G. Box, W. Hill, Discrimination among mechanistic models, Technometrics 9 (1) (1967) 57–71.
[10] E. Bradley, M. Easley, R. Stolle, Reasoning about nonlinear system identification, Artificial Intelligence 133 (1–2) (2001) 139–188.
[11] D. Burmaster, The unsteady continuous culture of phosphate-limited *Monochrysis lutheri Droop*: Experimental and theoretical analysis, J. Experimental Marine Biology and Ecology 39 (2) (1979) 167–186.
[12] J. Caperon, J. Meyer, Nitrogen-limited growth of marine phytoplankton I: Changes in population characteristics with steady-state growth rate, Deep-Sea Research 19 (1972) 601–618.
[13] B. Chen, S. Asprey, On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations, Industrial and Engineering Chemistry Research 42 (7) (2003) 1379–1390.
[14] D. Contois, Kinetics of bacterial growth: Relationship between population density and specific growth rate of continuous cultures, J. General Microbiology 21 (1959) 40–50.
[15] M. Cooney, K. McDonald, Optimal dynamic experiments for bioreactor model discrimination, Applied Microbiology and Biotechnology 43 (5) (1995) 826–837.
[16] C. Dariva, J. Oliveira, J. Pinto, Experimental design for model discrimination of thermodynamic models, Fluid Phase Equilibria 146 (1–2) (1998) 35–50.
[17] H. de Jong, Qualitative simulation and related approaches for the analysis of dynamical systems, Knowledge Engineering Review 19 (2) (2005) 93–132.
[18] H. de Jong, N.J.I. Mars, P. van der Vet, Computer-supported resolution of measurement conflicts: A case-study in materials science, Foundations of Science 4 (4) (1999) 427–461.
[19] H. de Jong, A. Rip, The computer revolution in science: steps towards the realization of computer-supported discovery environments, Artificial Intelligence 91 (2) (1997) 225–256.

[20] J. de Kleer, Using crude probability estimates to guide diagnosis, Artificial Intelligence 45 (3) (1990) 381–391.
[21] J. de Kleer, O. Raiman, M. Shirley, One step lookahead is pretty good, in: W. Hamscher, L. Console, J. de Kleer (Eds.), Readings in Model-based Diagnosis, Morgan Kaufmann, San Mateo, CA, 1992, pp. 138–142.
[22] J. de Kleer, B. Williams, Diagnosing multiple faults, Artificial Intelligence 32 (1) (1987) 97–130.
[23] M. Droop, Vitamin B12 and marine ecology IV: The kinetics of uptake growth and inhibition in *Monochrysis lutheri*, J. Marine Biological Assoc. 48 (3) (1968) 689–733.
[24] D. Dvorak, B. Kuipers, Process monitoring and diagnosis, a model-based approach, IEEE Expert 6 (2) (1991) 67–74.
[25] D. Espie, S. Macchietto, The optimal design of dynamic experiments, AIChE 35 (2) (1989) 223–229.
[26] B. Faltings, P. Struss (Eds.), Recent Advances in Qualitative Physics, MIT Press, Cambridge, MA, 1992.
[27] A. Farquhar, B. Kuipers, J. Rickel, D. Throop, Q.R. Group, QSIM: The program and its use, Technical report UT-AI-90-123, University of Texas, Austin, TX, 1993.
[28] V. Fedorov, Theory of Optimal Experiments, Academic Press, New York, 1972.
[29] V. Fedorov, P. Hackl, Model-Oriented Design of Experiments, Springer, New York, 1997.
[30] G. Gorry, G. Barnett, Experience with a model of sequential diagnosis, Computers and Biomedical Research 1 (1968) 490–507.
[31] W. Hamscher, L. Console, J. de Kleer (Eds.), Readings in Model-Based Diagnosis, Morgan Kaufmann, San Mateo, CA, 1992.
[32] G. Hardy, J. Littlewood, G. Pólya, Inequalities, Cambridge University Press, Cambridge, MA, 1967.
[33] G. Harris, Phytoplankton Ecology Structure, Function and Fluctuation, London, New York, 1986.
[34] R. Horst, P. Pardalos (Eds.), Handbook of Global Optimization, Kluwer, Dordrecht, 1995.
[35] T. Hsiang, P. Reilly, A practical method for discriminating among mechanistic models, Canadian J. Chemical Engineering 49 (1971) 865–871.
[36] S. Hsu, S. Hubbell, P. Waltman, A mathematical theory for single-nutrient competition in continuous cultures of microorganisms, SIAM J. Appl. Math. 32 (2) (1977) 366–383.
[37] T. Ideker, V. Thorsson, R. Karp, Discovery of regulatory interactions through perturbation: Inference and experimental design, in: R. Altman, K. Lauderdale, A. Dunker, L. Hunter, T. Klein (Eds.), Proceedings of the Pacific Symposium on Biocomputing, PSB, 2000, vol. 5, World Scientific Publishing, Singapore, 2000, pp. 302–313.
[38] H. Kay, SQSIM: A simulator for imprecise ODE models, Computers and Chemical Engineering 23 (1) (1998) 27–46.
[39] H. Kay, B. Rinner, B. Kuipers, Semi-quantitative system identification, Artificial Intelligence 119 (1–2) (2000) 103–140.
[40] J. Keppens, Q. Shen, On compositional modelling, Knowledge Engineering Review 16 (2) (2001) 157–200.
[41] B. Kuipers, Qualitative simulation: Then and now, Artificial Intelligence 59 (1–2) (1993) 133–140.
[42] B. Kuipers, Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge, MIT Press, Cambridge, MA, 1994.
[43] B. Kuipers, D. Berleant, Using incomplete quantitative knowledge in qualitative reasoning, in: Proceedings of the 7th National Conference on Artificial Intelligence, AAAI-88, Morgan Kaufmann, Los Altos, CA, 1988, pp. 324–329.
[44] R. Leitch, Q. Shen, Prioritising behaviors in qualitative simulation, in: J. McDermott (Ed.), Proceedings of the 13th International Joint Conference on Artificial Intelligence, IJCAI-93, Morgan Kaufmann, San Mateo, CA, 1993, pp. 1523–1528.
[45] C. Lund, C. Surko, M. Maple, S. Yamamoto, Model discrimination in oscillatory CO oxidation on platinum catalysts at atmospheric pressure, Surface Science 459 (2000) 413–425.
[46] S. McIlraith, R. Reiter, On tests for hypothetical reasoning, in: Readings in Model-Based Diagnosis, Morgan Kaufmann, San Mateo, CA, 1992, pp. 89–96.
[47] T. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.
[48] J. Monod, Recherches sur la croissance des cultures bactériennes, Hermann, Paris, 1942.
[49] S. Narasimhan, P. Mosterman, 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, Cape Cod, MA, 1998, pp. 94–101.
[50] S. Rajamoney, The design of discrimination experiments, Machine Learning 12 (1993) 185–203.
[51] C. Schut, B. Bredeweg, An overview of approaches to qualitative model construction, Knowledge Engineering Review 11 (1) (1996) 1–25.
[52] P. Struss, Mathematical aspects of qualitative reasoning, Artificial Intelligence in Engineering 3 (3) (1988) 156–169.
[53] 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.
[54] P. Struss, Fundamentals of model-based diagnosis of dynamic systems, in: M. Pollack (Ed.), Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, IJCAI-97, Morgan Kaufmann, San Francisco, CA, 1997, pp. 480–485.
[55] R. Takors, W. Wiechert, D. Weuster-Botz, Experimental design for the identification of macrokinetic models and model discrimination, Biotechnology and Bioengineering 56 (5) (1997) 564–576.
[56] L. Travé-Massuyès, P. Dague (Eds.), Modèles et raisonnements qualitatifs, Hermès, Paris, 2003.
[57] G. Treitz, G. Maria, F. Giffhorn, E. Heinzle, Kinetic model discrimination via step-by-step experimental and computational procedure in the enzymatic oxidation of D-glucose, J. Biotechnology 85 (3) (2001) 271–287.
[58] I. Vatcheva, Computer-supported experiment selection for model discrimination, Ph.D. thesis, University of Twente, Enschede, the Netherlands, 2001.
[59] I. Vatcheva, O. Bernard, H. de Jong, J.-L. Gouzé, N. Mars, Discrimination of semi-quantitative models by experiment selection: Method and application in population biology, in: B. Nebel (Ed.), Proceedings of the 17th International Joint Conference on Artificial Intelligence, IJCAI-01, Morgan Kaufmann, San Mateo, CA, 2001, pp. 74–79.
[60] I. Vatcheva, H. de Jong, Semi-quantitative comparative analysis, in: T. Dean (Ed.), Proceedings of the 16th International Joint Conference on Artificial Intelligence, IJCAI-99, Morgan Kaufmann, San Francisco, CA, 1999, pp. 1034–1040.

[61] I. Vatcheva, H. de Jong, N.J.I. Mars, Selection of perturbation experiments for model discrimination, in: W. Horn (Ed.), Proceedings of the 14th European Conference on Artificial Intelligence, ECAI-2000, IOS Press, Amsterdam, 2000, pp. 191–195.
[62] D. Weld, J. de Kleer (Eds.), Readings in Qualitative Reasoning about Physical Systems, Morgan Kaufmann, San Mateo, CA, 1990.
[63] S. Xia, N. Smith, Automated modelling: A discussion and review, Knowledge Engineering Review 11 (2) (1996) 137–160.