Selection of Perturbation Experiments for Model Discrimination

Ivayla Vatcheva[1], **Hidde de Jong**[2] and **Nicolaas J.I. Mars**[1]

Abstract. It often occurs that a system can be described by several competing models. In order to distinguish among the alternative models, further information about the behavior of the system is required. One way to obtain such information is to perform suitably chosen perturbation experiments. We introduce a method for the selection of optimal perturbation experiments for discrimination among a set of dynamical models. The models are assumed to have the form of semi-quantitative differential equations. The method employs an optimization criterion based on the entropy measure of information.

1 Introduction

Scientists and engineers are frequently faced with situations in which a given system can be described by several *competing models*. When analyzing the synthesis rate of a product in a catalized chemical reaction, one may end up with several equations that all satisfy a set of measurements [13]. For the mitotic clock of early embryos, a dozen of models predicting the observed periodic behavior of the concentrations of key proteins has been suggested [9].

In order to identify which of the given models best describes the actual situation, new observations have to be made. These can be obtained by performing supplementary *perturbation experiments*. In a perturbation experiment the structure of the system and/or the experimental conditions are changed. An experiment discriminates between the competing models, if the predictions of some of the candidates, which have been properly modified to reflect the experimental change, fit the newly obtained data whereas others do not. The problem of model discrimination can then be defined as the problem of selecting an experiment that gives rise to observations matching the predictions of as few of the models as possible.

Incomplete knowledge about the system to be understood does not always permit the formulation of detailed quantitative models. In what follows we assume that the models are given in the form of semi-quantitative differential equations. Appropriate semi-quantitative simulation techniques are used to derive interval predictions for the model variables. As measurements obtained in the experiments may be imprecise, they are considered to be intervals as well.

We present a method for the systematic choice of perturbation experiments for the discrimination of semi-quantitative dynamical models. Experiments are selected on the basis of a generalization of an entropy criterion suggested by Box & Hill [2], which measures the information increment provided by each experiment. The concept of entropy as a discrimination criterion has also been used in statistics (e.g. [10]), and in model-based diagnosis (e.g. [4, 12]). A novel aspect of our work is that we extend this concept to the case of perturbation experiments and to situations in which experimental systems are described by semi-quantitative dynamical models.

The in-principle applicability of our approach is illustrated on a set of competing models of an oscillatory, second-order system. We will consider six models of a mass-spring system and illustrate the choice of suitable perturbations to discriminate between the models. The principles involved in this example are applicable to the investigation of more complex and less understood oscillating systems.

The presentation starts with a description of the problem of model discrimination. A number of basic concepts are introduced and the relationship between models and experiments is given. The criterion for choosing a maximally-discriminating perturbation is described in Sec. 3. In Sec. 4, the application of the method is illustrated on the example. Sec. 5 discusses limitations and extensions of our method, in the context of related work in statistics and model-based diagnosis.

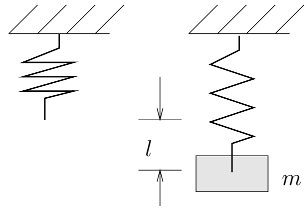
2 Model discrimination by perturbation experiments

The systems we will be concerned with in this paper are (physical) systems controlled in experiments, also called *experimental systems*. An example of an experimental systems is a cell culture allowed to grow under controlled environmental conditions, including nutrient supply and temperature. Control over an experimental system is achieved by creating and maintaining its structure and by regulating the experimental conditions under which the behavior of the system evolves.

Suppose a set M of models of an experimental system has been proposed. Let $p(m_i)$ be the a priori probability of $m_i \in M$ being the correct model of the system. The probabilities can be derived from preliminary observations on the system behavior or theoretical considerations. If no prior knowledge about the relative plausibilities of the models exists, equal probabilities are assumed. We say that the models in M are *competing*. M is assumed to be complete, that is, $\sum_{m_i \in M} p(m_i) = 1$.

In this paper we will model experimental systems by means of *semi-quantitative differential equations (SQDEs)*, that is, qualitative differential equations (QDEs) enhanced with numerical information. The quantitative information completing a QDE takes the form of numerical ranges added to landmarks and of envelopes for monotonic function constraints [1]. Fig. 1 shows two SQDEs describing a simple mass-spring system. The models assume that the forces playing a role in the experiment are a spring force and a friction force, but they differ in the precise nature ascribed to the former.

[1] Department of Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands, email: {ivayla,mars}@cs.utwente.nl
[2] Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de Recherche Rhône-Alpes, 655 avenue de l'Europe, 38330 Montbonnot Saint-Martin, France, email: Hidde.de-Jong@inrialpes.fr



$$
\begin{aligned}
(m_1)\ QV(\dot{x}) &= QV(v)\\
QV(\dot{v}) &= QV(a)\\
QV(a) &= -\frac{QV(g)}{QV(l)}QV(x)\\
&\quad -\frac{QV(c)}{QV(m)}QV(v)
\end{aligned}
\qquad
\begin{aligned}
(m_2)\ QV(\dot{x}) &= QV(v)\\
QV(\dot{v}) &= QV(a)\\
QV(a) &= -\frac{QV(g)}{QV(l)}\left(QV(x)-\frac{QV(x^3)}{QV(k)}\right)\\
&\quad -\frac{QV(c)}{QV(m)}QV(v)
\end{aligned}
$$

Figure 1. Damped mass-spring system and two QDEs m_1 and m_2 describing the system (for the notation, see [14]). The variables refer to the position x, velocity v, acceleration a, mass m, gravity constant g, initial spring elongation l and friction constant c. The following intervals complete the QDEs to SQDEs: $range(m) \in [2.95, 3.05]$, $range(g) = [9.83, 9.83]$, $range(l) = [5.8, 6.0]$, $range(c) = [0.3, 0.4]$. The initial values for the position and the velocity are $[0.9, 1.1]$ and $[0, 0]$, respectively. The constant k is specified by $range(k) = [6, 6]$.

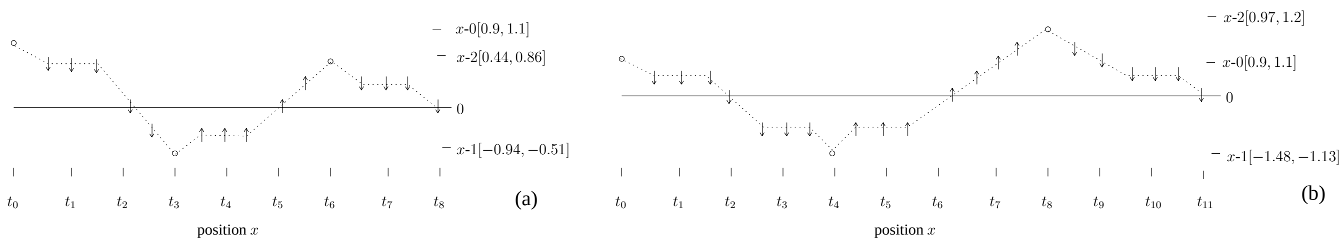


Figure 2. SQBs derived from m_1 (Fig. 1). The SQB in (b) is obtained after a perturbation of m_1, increasing the initial velocity to $[1.9, 2.1]$.

To distinguish between the models, additional information about the system is required. This information can be obtained by performing a suitably chosen *perturbation experiment*. In a perturbation experiment the system structure or the experimental conditions are modified. The changes have to be reflected on the competing models in such a way that the operations on a model correspond with perturbations of the experimental system (Fig. 3).

$$
\begin{array}{ccccc}
\text{model} & \xrightarrow{\text{perturbation}} & \text{perturbed model} & \xrightarrow{\text{prediction}} & \text{predictions}\\
\text{experimental system} & \xrightarrow{\text{perturbation}} & \text{perturbed experimental system} & \xrightarrow{\text{execution of experiment}} & \text{observations}
\end{array}
$$

Figure 3. Correspondence between perturbation experiments and model perturbations.

In order to predict the consequences of a perturbation, we employ the semi-quantitative simulation techniques Q2 and Q3 [1, 8]. Q2 and Q3 exploit the semi-quantitative information in an SQDE to refine a qualitative behavior tree produced by QSIM. More specifically, they rule out qualitative behaviors or transform qualitative behaviors into *semi-quantitative behaviors (SQBs)* in which the qualitative values are annotated with numerical ranges. Fig. 2 shows the oscillations predicted by model m_1, before and after a perturbation that consists in releasing the object with nonzero initial velocity.

A semi-quantitative behavior is a prediction of the interval value of the variables at the *distinguished time-points*, the time-points at which some variable changes its qualitative value. For instance, the SQB in Fig. 2(a) shows that at t_3, the time-point at which x reaches its maximum for the first time, the value of x lies in the interval $[-0.94, -0.51]$.

In addition to predictions of the value of a variable at a time-point, we might be interested in the difference in value of a variable before and after a perturbation. Predictions of the relative interval value of variables can be obtained by subtracting the predictions at corresponding distinguished time-points in the behavior before and after a perturbation, so-called *meaningful pairs of comparison* [3]. As a consequence of the use of semi-quantitative information, these predictions may be weaker than necessary. We use the comparative analysis technique SQCA to obtain more precise predictions [14]. The information in Fig. 2 allows one to infer, by subtracting ranges, that the difference $\hat{x} - x$ at the pair of comparison $\langle t_3, \hat{t}_4 \rangle$ lies in the interval $[-0.97, -0.19]$. $\hat{x}$ and $\hat{t}$ refer to variables in the perturbed system. Application of SQCA refines this prediction by narrowing the interval to $[-0.93, -0.19]$.

The amplitude in the behavior of the perturbed mass-spring system, or the difference in amplitude in the behaviors of the perturbed and unperturbed systems, are examples of *behavioral features* that help in discriminating competing models of a system. A predicted behavioral feature is an interval value calculated from a set of predictions by means of an arithmetic function. This may simply be a predicted value or a relative value, as in the case of a predicted amplitude. A less trivial feature is the frequency of an oscillation, which can be calculated from the interval ranges of the distinguished time-points of two successive maxima.

Predicted behavioral features need to correspond with *observed* behavioral features of the system. That is, it should be possible to relate a predicted behavioral feature to some direct or indirect measurement of quantities of the system. As measurements will be assumed to have the form of *confidence intervals*, observed behavioral features are intervals.

The results of a perturbation experiment can be used to recompute the probabilities of the competing models. Models of which the predictions do not agree with the observations will have an a posteriori probability equal to 0. The model discrimination problem can now be intuitively stated as follows: find the experiment with values for the observed behavioral features that make a maximum number of models improbable. In the next section, we elaborate this intuition by means of an approach based on concepts from information theory.

3 Method for the selection of perturbation experiments

We will be interested in finding the perturbation yielding the highest increment in information [2]. Consider a behavioral feature Y, with interval values in $D \subseteq \mathbb{R}$. Let $e \in E$ be a perturbation experiment, whose outcome yields a value $Y^e = [y^e - \epsilon/2, y^e + \epsilon/2]$ of the behavioral feature, where y^e is the midpoint of the interval Y^e and ϵ is the size of the confidence interval for Y. The *information increment* of e is formulated as

$$\Delta H(e) = -\sum_{m_i \in M} p(m_i) \ln p(m_i) + \sum_{m_i \in M} p(m_i \mid Y^e) \ln p(m_i \mid Y^e), \quad (1)$$

where $p(m_i)$ and $p(m_i \mid Y^e)$ are the a priori and a posteriori probabilities of m_i. ΔH reaches its maximum when the a posteriori probabilities of all models but one are 0. A minimal value is attained when the a posteriori probabilities are equal.

The $p(m_i \mid Y^e)$s in (1) are not known, since they are determined by the outcome of the experiment. However, we can express the expected value of ΔH in terms of the probability distributions $g_i^{\{e,Y\}}$ of the behavioral feature Y. For brevity, g_i^e instead of $g_i^{\{e,Y\}}$ will be used if no confusion about the behavioral feature being considered is possible. The value of Y predicted by m_i under perturbation e is an interval $V_i^e \subseteq D$, with distribution $g_i^e : D \rightarrow \mathbb{R}_{\geq 0}$ defined as follows

$$g_i^e(y) = \begin{cases} \dfrac{|[y - \frac{\epsilon}{2}, y + \frac{\epsilon}{2}] \cap V_i^e|}{|V_i^e|} & , [y - \frac{\epsilon}{2}, y + \frac{\epsilon}{2}] \cap V_i^e \neq \emptyset, \\ 0 & , [y - \frac{\epsilon}{2}, y + \frac{\epsilon}{2}] \cap V_i^e = \emptyset, \end{cases} \quad (2)$$

where $|\cdot|$ denotes an interval length. $g_i^e(y)$ expresses the probability that the empirically-determined value of Y is $[y - \epsilon/2, y + \epsilon/2]$ if model m_i is the correct model. (2) can be replaced by the following equivalent expression, where the g_i^es are defined as piecewise-linear functions:

$$g_i^e(y) = \begin{cases} \dfrac{y - \underline{V_i^e} + \epsilon/2}{|V_i^e|} & , y \in [\underline{V_i^e} - \epsilon/2, \underline{V_i^e} + \epsilon/2], \\ \dfrac{\epsilon}{|V_i^e|} & , y \in [\underline{V_i^e} + \epsilon/2, \overline{V_i^e} - \epsilon/2], \\ \dfrac{-y + \overline{V_i^e} + \epsilon/2}{|V_i^e|} & , y \in [\overline{V_i^e} - \epsilon/2, \overline{V_i^e} + \epsilon/2], \\ 0 & , y \notin [\underline{V_i^e} - \epsilon/2, \overline{V_i^e} + \epsilon/2], \end{cases} \quad (3)$$

and $\underline{V_i^e}$ and $\overline{V_i^e}$ denote the lower and the upper bound of V_i^e, respectively. Fig. 4 illustrates the function g_3^e for an experiment e consisting of replacing the object in the mass-spring system by a lighter object. The behavioral feature in this case is the interval value for the amplitude of the system.

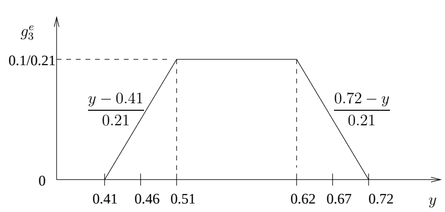


Figure 4. A plot of the function g_3^e for the amplitude of the mass-spring system in an experiment consisting of replacing the object by a lighter object (e_4 in the next section). The prediction of model m_3 (Fig. 6) perturbed according to this experiment is $V_3^e = [0.46, 0.67]$, and $\epsilon = 0.1$.

Call the expected value of the information increment $\Delta J(e)$. By definition,

$$\Delta J(e) = \int_{y \in D} \Delta H(e) g^e(y) dy, \quad (4)$$

where $g^e(y) = \sum_{m_i \in M} p(m_i) g_i^e(y)$. By substituting the expression for $\Delta H(e)$ in (4) we get,

$$\Delta J(e) = \sum_{m_i \in M} p(m_i) \int_{y \in D} g_i^e(y) \{ \sum_{m_j \in M} p(m_j \mid Y) \ln p(m_j \mid Y) - \sum_{m_j \in M} p(m_j) \ln p(m_j) \} dy, \quad (5)$$

where $Y = [y - \epsilon/2, y + \epsilon/2]$ and

$$p(m_j \mid Y) = \frac{p(m_j) g_j^e(y)}{g^e(y)} \quad (6)$$

via the Bayes rule. Combination of (5) and (6) gives, after algebraic simplification,

$$\Delta J(e) = \sum_{m_i \in M} p(m_i) \int_{y \in D} g_i^e(y) \ln \frac{g_i^e(y)}{g^e(y)} dy. \quad (7)$$

The optimal next experiment to perform is the one for which (7) is maximized. Intuitively, the criterion favors experiments for which the corresponding model perturbation results in predicted intervals for the behavioral feature that overlap as little as possible. On average, less overlap of the intervals will increase the chance that a measurement of the feature discriminates between the models. This can be illustrated by means of the predictions of the relative amplitude by three alternative, equiprobable models of the mass-spring system. Consider the case of a perturbation e_1 replacing the medium with an almost frictionless medium (setting c to 0), and a perturbation e_3 increasing the mass to $[11.95, 12.05]$ (see Fig. 5). The expected information increment is higher for e_3, as the predicted intervals have less overlap.

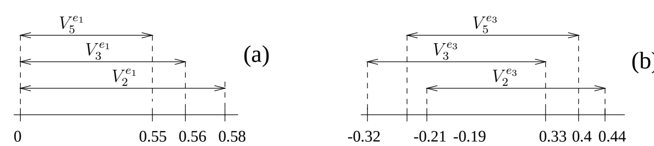


Figure 5. Two sets of predicted behavioral features (relative interval values for the amplitude). In (a) $V_2^{e_1}, V_3^{e_1}$ and $V_5^{e_1}$ are obtained from m_2, m_3 and m_5, with perturbations according to e_1 (see next section). In (b) $V_2^{e_3}, V_3^{e_3}$ and $V_5^{e_3}$ are obtained from the same models with perturbations according to e_3. $\Delta J(e_1) = 0.0176$ in (a), and $\Delta J(e_3) = 0.1246$ in (b).

If several behavioral features $Y_1, \ldots, Y_k$ are taken into account, the formula in (7) remains unchanged, except for replacing y by $\boldsymbol{y}$, D by $\boldsymbol{D} = D_1 \times \ldots D_k$, the distributions $g_i^e(y)$ by joint probability distributions $g_i^{\{e, Y_1, \ldots, Y_k\}}(\boldsymbol{y})$, and the integral by a multiple integral. Intuitively, the criterion now tries to maximize the non-overlapping parts of the k-dimensional boxes in $\boldsymbol{D}$ that are defined by the values for the behavioral features predicted by the m_is.

The soundness of the simulation algorithms referred to above guarantees that the models will never be falsely discriminated. If the measurement of the corresponding observed behavioral feature is correct, this implies that a model will never be rejected on false grounds. However, as a consequence of the incompleteness of the algorithms, competing models may fail to be discriminated while they should be.

On the basis of the selection criterion, a simple algorithm can be imagined to identify the model from M (if any) that best describes the real system by means of a minimal number of experiments. Let θ be a number between 0 and 1, determining the threshold above which we consider a model to be the best representation of the system. That is, m_i is assumed to best describe the system if $p(m_i) \geq \theta$. Let

$$QV(a) = -\frac{QV(g)}{QV(l)} QV(x) - \frac{QV(c)}{QV(m)} QV(|v|) QV(v)$$

(m_3)

$$QV(a) = -\frac{QV(g)}{QV(l)} \left(QV(x) - \frac{QV(x^3)}{QV(k)} \right) - \frac{QV(c)}{QV(m)} QV(|v|) QV(v)$$

(m_4)

$$QV(a) = -\frac{QV(g)}{QV(l)} \left(QV(x) + \frac{QV(x^3)}{QV(k)} \right) - \frac{QV(c)}{QV(m)} QV(v)$$

(m_5)

$$QV(a) = -\frac{QV(g)}{QV(l)} \left(QV(x) + \frac{QV(x^3)}{QV(k)} \right) - \frac{QV(c)}{QV(m)} QV(|v|) QV(v)$$

(m_6)

Figure 6. Models m_3-m_6 together with m_1 and m_2 in Fig. 1 form a set of competing models for the mass-spring system. Models m_1 and m_3 assume linear spring force. Models m_2 and m_4 assume soft spring forces (the stiffness of the spring decreases with the displacement), while the spring force in m_5 and m_6 is hard (the stiffness increases with the displacement). In m_1, m_2 and m_5 the acceleration depends linearly on the velocity. The models m_3, m_4 and m_6 assume quadratic dependency. The meaning of the variables and the constraints for $QV(\dot{x})$ and $QV(\dot{v})$ is the same as in Fig. 1.

(a)

	e_1	e_2	e_3	e_4	e_5
m_1	[0.9, 1.1]	[0.12, 0.39]	[0.8, 1.01]	[0.63, 0.97]	[1.13, 1.48]
m_2	[0.9, 1.1]	[0.11, 0.38]	[0.81, 1.01]	[0.62, 0.88]	[1.57, 2.02]
m_3	[0.9, 1.1]	[0.29, 0.52]	[0.88, 1.01]	[0.46, 0.67]	[1.24, 1.65]
m_4	[0.9, 1.1]	[0.34, 0.55]	[0.8, 1.0]	[0.41, 0.62]	[1.22, 1.64]
m_5	[0.9, 1.1]	[0.18, 0.43]	[0.81, 1.01]	[0.64, 0.86]	[1.34, 1.77]
m_6	[0.9, 1.1]	[0.23, 0.51]	[0.83, 1.02]	[0.41, 0.62]	[1.09, 1.47]

(b)

	$\Delta J(e_i)$
e_1	0
e_2	0.5315
e_3	0.0934
e_4	0.7499
e_5	0.6504

(c)

	$\Delta J(e_i)$
e_1	0
e_2	0.3163
e_3	0.1239
e_5	0.1694

Table 1. (a) Predictions for feature f_1, the interval value for the amplitude derived from the models for the perturbations $e_1, \cdots, e_5$. (b) The values of ΔJ computed for all perturbations, and (c) some values of ΔJ after application of e_4 (see text).

$p(m_j)$ be the a priori probabilities of the models and E a set of predefined perturbation experiments.

while ($\exists m_i \in M : p(m_i) \neq 0$ **and** $\forall m_i \in M : p(m_i) < \theta$
and not E empty) **do**
determine $e \in E$ for which $\Delta J(e)$ is maximal
perform experiment corresponding to e, determine Y^e
compute the a posteriori probabilities $p(m_j|Y^e)$
set $p(m_j)$ to $p(m_j|Y^e)$
remove e from E

The algorithm selects perturbation experiments until one of the following happens: a model has a sufficiently high probability, all models have zero probabilities, or all possible experiments have been executed. If the algorithm terminates with $p(m_i) = 0$ for all models, obviously the assumption for completeness of M is violated.

4 Example and evaluation

Consider the six models of a mass-spring system listed in Fig. 1 and Fig. 6 [11]. The models differ in the terms for the spring and the friction force. The experiment consists in stretching and then releasing the spring. Assume the following perturbation experiments can be performed. e_1: replace the medium by an approximately frictionless medium ($c = 0$); e_2: replace the medium by a more compact medium ($c = [2.85, 3.15]$); e_3: test with a heavier object having mass $[11.95, 12.05]$; e_4: test with a lighter object having mass $[0.7, 0.8]$; e_5: release the object with initial velocity $[1.9, 2.2]$.

We consider four behavioral features: Y_1 is the interval value of the maximum distance from the rest position (the amplitude); Y_2 is the interval value of the frequency of the system; Y_3 is the relative interval value of the maximum amplitude for the perturbed and the original system; and Y_4 is the relative interval value of the frequency.

Values for Y_1 to Y_4 have been derived from the perturbed models by means of semi-quantitative simulation and comparative analysis. The predicted intervals for the amplitude are shown in Table 1(a). The first perturbation gives rise to identical predictions from all models. It is evident, even without looking at the value of $\Delta J(e_1)$, that the corresponding experiment will never distinguish between the models. The rest of the perturbations also do not give distinct intervals for this feature, but the predictions are not entirely overlapping. Hence, the measurements in the corresponding experiments may discriminate between at least some of the models.

Assume the amplitude of the system to be the only quantity measured in the experiments. Suppose the models have equal a priori probabilities $p(m_1) = \ldots = p(m_6) = 1/6$ and $\theta = 0.75$, that is, a model is considered best if its probability is larger than 0.75. At the first step of the algorithm, e_4 is chosen since it maximizes ΔJ (see Table 1(b)). Assume the experiment is executed and a measurement $[0.4207, 0.5207]$ is obtained. The measurement is not consistent with the predictions derived from m_1, m_2, and m_5 for this perturbation, so that the a posteriori probabilities of these models become 0. The a posteriori probabilities of the other three models after the experiment are $p(m_3) = 0.2330$, $p(m_4) = 0.3835$ and $p(m_6) = 0.3835$. In the next iteration, e_2 is selected (Table 1(c)). Assume the measurement $[0.3080, 0.4080]$ is obtained which gives rise to the posterior probabilities $p(m_3) = 0.2794$, $p(m_4) = 0.3428$ and $p(m_6) = 0.3778$. Next, e_5 is chosen. A measurement $[1.1340, 1.2340]$ causes $p(m_3) = 0$ and the algorithm terminates, giving m_6 as the best model of the system with $p(m_6) = 0.8964$.

In order to evaluate the performance of the method we have adopted the following strategy. First, one of the models (m_6) was arbitrary selected. "Experimental" data was then produced by generating random intervals within the predictions of m_6. The length of the random intervals was set equal to the size of the confidence interval of the behavioral feature ($\epsilon = 0.1$ in the case of the amplitude). Finally, the algorithm of the previous section was applied given these data. This procedure was repeated 20 times and the results analyzed.

In only 15% of the cases the model that was used to generate the data was identified as the single remaining candidate. In the rest of the cases the algorithm terminated with two to three candidate models that could not be discriminated. On average, for the identification of the model 4 experiments were necessary. For comparison, when the size of the confidence interval was taken to be 0.01, in 40% of the cases m_6 was identified with average number of experiments 2.5. The results show, not surprisingly, that when the measurement error is smaller, better discrimination is achieved.

Now suppose all four features are considered, the other circumstances remaining the same. In this case, e_5 maximizes ΔJ and it is selected as the best experiment (see the table below). Values of $[1.134, 1.234]$ and $[4.12, 4.22]$ for the amplitude and the period of the perturbed system, for instance, give rise to the posteriori probabilities $p(m_1) = \ldots = p(m_5) = 0$ and $p(m_6) = 1.0$.

	e_1	e_2	e_3	e_4	e_5
$\Delta J(e_i)$	0.4049	1.1429	1.5548	1.5285	3.1811

The above evaluation procedure was again applied 20 times, now for the situation that all four features are taken into account. We found that the average number of experiments necessary to identify model m_6 was 1.1. In only two of the cases a second iteration in the algorithm was necessary. In all cases complete discrimination was achieved.

The example illustrates that when more behavioral features are considered, a higher efficiency may be achieved: measuring only the amplitude, we needed four experiments to discriminate between the models, while taking into account all four behavioral features a single experiment turned out to be sufficient.

Evaluation by means of random data was used to investigate the performance improvement of the algorithm for experiment selection with respect to random selection of perturbation experiments. Assume the amplitude of the system is the only quantity being measured ($\epsilon = 0.1$). After 20 times we again obtained that 4 experiments are necessary, on average, to identify the correct model. The reasons for the lack of improvement of our method with respect to random selection are the large overlap between (some of) the predictions, the high measurement error assumed, and the low number of experiments provided. However, selecting the experiments in random order when all four features were considered, required 3.2 experiments on average to identify the correct model, whereas selecting the experiments by our method required only 1.1 experiments.

5 Discussion and conclusion

We have presented a method for discriminating among competing semi-quantitative models by selecting suitable perturbation experiments. The method chooses a maximally-discriminating experiment by means of a criterion based on the entropy measure of information. The application of this criterion was illustrated in an example concerning a set of competing models of a mass-spring system.

Information theory has been used in model-based diagnosis to distinguish among competing diagnoses of a faulty system (e.g. [4]). Like our method, these methods proceed by making new observations on the system. However, the work mentioned above is limited to determining the best measurement point within a given experiment, while we seek the best experiment that would permit optimal discrimination. Struss [12] has extended the approach in [4] by finding the best operating conditions that would give rise to the most discriminatory observations. Our work attempts to generalize this method by employing dynamical models and by extending the concept of discriminating test to discriminating perturbation experiment.

In statistics, the idea of employing the entropy measure as a discrimination criterion has been illustrated for distinguishing between quantitative algebraic models (e.g. [2, 10]). In [7] the entropy has been used to design observations discriminating among rival water quality models. However, these examples are restricted to fully numerical models with precise point measurements. In this paper, we have shown how the criterion can be generalized to the case that only imprecise, approximate observations of the system are available.

The idea of planning perturbation experiments for model discrimination based on an entropy measure has also been proposed in [6]. Models of a genetic regulation network are discriminated by varying the expression level of involved genes or the influence of external stimuli. This method, however, is limited to models in the form of Boolean networks and to binary perturbations. This article generalizes the approach in [6] by employing more advanced dynamical models and by extending the concept of perturbation experiments.

The work presented here can be extended into several directions. In practice, the number of possible perturbations will be infinite when the value of a quantity can be changed continuously. The problem of model discrimination as defined here should then be generalized. Instead of selecting a discrete perturbation that has been specified beforehand, a value for the quantity that maximizes (7) has to be chosen. An issue neglected thus far are the costs associated with experiments. In practice, the costs for performing an experiment may need to be balanced against its expected utility. In these cases, the problem can be reformulated as the selection of an experiment that maximizes $\Delta J(e)/h(cost(e))$, where h is a function depending on the intended application: one may be interested in effective experiments without caring about expenses, or prefer less costly tests.

Further research will concentrate on the extension of the method along the lines mentioned above, its comparison with other model discrimination techniques (e.g. [5]), and its application to real-world systems. Currently, we are applying the approach to a model discrimination problem in biology: the regulation of the cell cycle in early embryos [9]. This system is described by second-order models and exhibits periodic behavior similar to the oscillations of the mass-spring example considered here.

REFERENCES

[1] D. Berleant and B. Kuipers, ‘Qualitative and quantitative simulation: Bridging the gap’, *Artif Intell*, **95**, 215–256, (1997).

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

[3] H. de Jong and F. van Raalte, ‘Comparative envisionment construction: A technique for the comparative analysis of dynamical systems’, *Artif Intell*, **115**, 145–214, (1999).

[4] J. de Kleer and B.C. Williams, ‘Diagnosing multiple faults’, *Artif Intell*, **32**, 97–130, (1987).

[5] T. Hsiang and P.M. Reilly, ‘A practical method for discriminating among mechanistic models’, *Can J Chem Eng*, **49**, 865–871, (1971).

[6] T.E. Ideker, V. Thorsson, and R.M. Karp, ‘Discovery of regulatory interactions through perturbation: inference and experimental design’, in *Proceedings of Pacif. Symp. Biocomp.*, 5, pp. 302–313, (2000).

[7] J. Kettunen, H. Sirvio, and O. Varis, ‘Design of observations for discrimination among rival water quality models’, in *Optimal Design and Analysis of Experiments*, eds., Y. Dodge, V.V. Fedorov, and H.P. Wynn, 257–268, Elsevier Science Publishers, (1988).

[8] B. Kuipers, *Qualitative Reasoning*, MIT Press, 1994.

[9] M.N. Obeyesekere, S.L. Tucker, and S.O. Zimmerman, ‘Mathematical models of the cellular concentrations of cyclin and MPF’, *Biochem Bioph Res Co*, **184**(2), 782–789, (1992).

[10] P.M. Reilly, ‘Statistical methods in model discrimination’, *Can J Chem Eng*, **48**, 168–173, (1970).

[11] J.J. Stoker, *Nonlinear vibrations in mechanical and electrical systems*, Wiley, 1992.

[12] P. Struss, ‘Testing for discrimination of diagnoses’, in *Working Notes DX-94*, (1994).

[13] H.M. Swaan, *The oxidative dehydrogenation of ethane using promoted lithium doped magnesium oxide catalysts*, Ph.D. dissertation, University of Twente, 1992.

[14] I. Vatcheva and H. de Jong, ‘Semi-quantitative comparative analysis’, in *Proceedings IJCAI’99*, pp. 1034–1040, (1999).