Structure Identification by Optimized Interventions

Alberto Giovanni Busetto* **Joachim M. Buhmann**
Department of Computer Science, ETH Zurich, Zurich, Switzerland
Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland
{busettoa,jbuhmann}@inf.ethz.ch

Abstract

We consider the problem of optimal experimental design in structure identification. Whereas standard approaches simply minimize Shannon's entropy of the estimated parameter posterior, we show how to select between alternative model configurations, too. Our method specifies the intervention that makes an experiment capable of determining whether or not a particular configuration hypothesis is correct. This is performed by a novel clustering technique in approximated Bayesian parameter estimation for non-linear dynamical systems. The computation of the perturbation that minimizes the effective number of clusters in the belief state is constrained by the increase of the expected Kullback-Leibler divergence between the parameter prior and the posterior. This enables the disambiguation of persisting alternative explanations in cases where standard design systematically fails. Its applicability is illustrated with a biochemical Goodwin model, showing correct identification between multiple kinetic structures. We expect that our approach will prove useful especially for complex structures with reduced observability and multimodal posteriors.

1 INTRODUCTION

Experimental design refers to the general task of specifying all aspects of an experiment. The designer chooses the values of the variables that can be controlled before or during the process and is interested in the effect of some interventions on the experimental outcome. A design is called optimal if it maximizes a specified information optimality criterion. Due to its generality, this framework has a broad spectrum of applications and its renewed interest is justified by its fundamental role in systems biology. Not only contemporary experiments are technically laborious and demanding, but also the studied phenomena are more and more complex, with interacting structures whose identification is challenging. Therefore, two main goals are particularly desirable: the reduction of the uncertainty for the estimated model parameters and, more importantly, the ability to identify the correct structure between alternative configurations.

This paper is structured as follows. Section 1 describes the problem and summarizes our contributions, accompanied by an outline of the related work. Section 2 presents the method and its properties. The effectiveness of the method is exemplified in Section 3 for a biochemical Goodwin model, where it enables the correct identification of the kinetic structure. Finally, conclusions and outlook are drawn in Section 4.

1.1 PROBLEM DESCRIPTION

In this paper, we address the following problem: given the available prior information about a dynamical system, find the intervention that makes an experiment capable of decisively determining whether or not a parameter structure hypothesis is correct. In our scenario, the parameters define the network topology of the direct interactions between the variables of the dynamical system. We want to enable the automatic active selection between alternative explanations that would otherwise persist. Two constraints are added to define a useful and realistic problem: the feasibility of the intervention and the increase of the expected Kullback-Leibler Divergence (KLD) between the parameter prior and the posterior.

*Life Science Zurich PhD Program on Systems Biology of Complex Diseases

Appearing in Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS) 2009, Clearwater Beach, Florida, USA. Volume 5 of JMLR: W&CP 5. Copyright 2009 by the authors.

To illustrate the problem, let us consider the following example. As portrayed in Fig. 1, the position of a stationary object is measured by an observer. Due to the presence of two perfect mirrors, the object appears on the image plane at positions $-A$ and A. There is no way, solely based on observations, to disambiguate the positions. In this example, our goal is the determination of an appropriate intervention u that modifies the placement of the object, so that the resulting projected images allow the identification of the position. Moreover, the intervention must be feasible and within a given energy cost. After the perturbation, the object appears on the image plane at position $-B$ and its reflected image at position $B \neq C$, thus permitting the unambiguous determination of the state.

Our work is motivated by the renewed interest in the identification of nonlinear dynamical systems with complex structure. In the belief state, parameter clusters encode alternative underlying structures with local uncertainty. In general, developing such a strategy is challenging for at least two reasons. First, the associated estimation method must avoid the arbitrary exclusion of candidate solutions. Second, the simple reduction of the uncertainty in the parameter estimation in not enough when alternative explanations persist under the form of multimodal parameter posteriors. In fact, a belief state which is almost reduced to a finite set of Dirac deltas has low entropy but still some uncertainty remains between the alternative parameter configurations. Viceversa, in a large number of applications it is more desirable to tolerate some additional local uncertainty in the parameters, but finding a single identifiable structure.

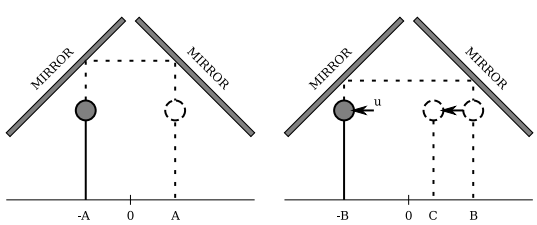

Figure 1: Illustration of the mirror example: an appropriate perturbation u allows the identification of the position of the object.

1.2 RELATED WORK

Experiment design has previously been conducted with special-case algorithms, often considering linear approximations (Seeger 2007). Recently, there has been a surge of interest in optimal design for systems biology and a large number of techniques have been proposed (Wu 2008). The vast majority of these addresses the problem of parameter estimation. Despite the existence of such tools and their intuitive usefulness, little is known about the limits of their ability to discriminate between alternative parameter configurations. This is mainly due to the intrinsic difficulty coming from the non-linearity of the studied dynamical systems (Akaho 2007). From the computational point of view, the stable detection of multiple configurations is already a challenging task. This is caused by the fast divergence in the sequential importance sampling/resampling approximations of Bayesian state estimation (King 2000).

In the case of Bayesian optimal experiment design, a common goal is the maximization of the expected KLD between the prior and the posterior. When posteriors are approximated by Gaussian distributions, this corresponds to the popular D-optimality criterion. There are theoretical and practical limitations for this approach, among others

- the Gaussian assumption do not hold for posteriors of non-linear systems and is a poor approximation especially for systems with complex and interesting dynamics,
- as described before, the maximization of the expected KLD between the prior and the posterior do not necessarily discriminate between alternative parameter structures.

1.3 CONTRIBUTIONS

We address the problem of structure identification by proposing a novel design approach that actively intervenes to generate crucial experiments. While reducing the uncertainty in the parameter estimation, our framework enables the ability to choose between mutually exclusive explanations of the observed phenomena. This feature overcomes some of the limitations of standard design described before thanks to a novel embedded clustering. We introduce a new resampling approach, which surmounts the common unimodal degeneracy problem of sequential Monte Carlo algorithms. The resampling is preventive and based on clustering: we show that a higher concentration of samples in the most significant regions of the parameter space maintains stability and facilitates convergence.

We combine clustering and sequential Monte Carlo state estimation in a new criterion, that allows us to automatically detect and actively reduce the number of modes in the belief state. We demonstrate that the sole maximization of the KLD is not always able to discriminate between alternative families of parameter configurations. This situation is influenced by the degree of completeness of the measurement process and by the complexity of the dynamic behavior. As an

additional advantage, our scheme can be applied to general nonlinear systems both for discrete and continuous time. Finally, we explore the effects of the energy cost function and of feasibility constraints for experimental interventions.

The results suggest that the stability introduced by the dynamic clustering has strong effects on the ability to discriminate between competing hypotheses, thus validating our framework as an appropriate design for systems with complex behavior.

2 METHODS

We begin with the mathematical specification of the problem, followed by a description of the method. Our approach is based on the generation of an equivalent extended system, which facilitates the design process. The latter is performed by minimizing the number of clusters in the predicted parameter posterior, under the constraints of feasibility and of increased expected KLD between consecutive belief states.

2.1 GENERAL EXTENDED MODEL

Let us start by considering the canonical process/measurement representation for a continuous-time continuous-valued stochastic dynamical system observed at discrete points in time. Let Σ be a system whose process model is given, in a time span $[t_0, t_{\text{end}}]$, by the Itō Stochastic Differential Equation (SDE)

$$d\mathbf{z}_t = f_z(\mathbf{z}_t, \theta, u_t, t)dt + \sigma_z(\mathbf{z}_t, u_t, t)dW_{zt}, \qquad (1)$$

where $\mathbf{z}_t \in \mathbb{R}^n$ is the state vector with finite dimension n, $\theta \in \mathbb{R}^r$ is the parameter vector and $u_t \in \mathcal{U}$ is a controllable input intervention. Note that f_z is a known time-dependent non-linear function that, given the current state, the parameters and the input, governs the deterministic component of the r.h.s. in (1). It represents the deterministic drift acting upon W_{zt}, that is a standard n-dimensional Wiener process with infinitesimal variance σ_z^2 in the infinitesimal time interval dt. It is now possible to consider an equivalent extended state space of dimension $n + r$, where the extended states $\mathbf{x}$ are given by a combined transformation of the original states and parameters θ. In this way, the parameters are hidden in the initial conditions $\mathbf{x}_0$ of the extended system. The parameter estimation task is transformed into an equivalent state estimation problem (Doucet 2003). Consequently, (1) can be rewritten as

$$d\mathbf{x}_t = f(\mathbf{x}_t, u_t, t)dt + \sigma(\mathbf{x}_t, u_t, t)dW_t, \qquad (2)$$

where f and σ are functions that satisfy the equivalence between (1) and (2) for the dynamics of the original process model. The same notation applies to the equivalent extended Wiener process W_t.

The measurement process is given by

$$\mathbf{y}_i = h(\mathbf{x}_{t_i}, V_t), \qquad (3)$$

where $\mathbf{y}_i \in \mathbb{R}^m$ is the experimental output, h is the known measurement function and V_t is a noise term. The measurement is performed only at discrete time points $t_i \in [t_0, t_{\text{end}}]$, that are arbitrarily distributed and sequentially indexed. The result of the experiment is given by the data set $\mathcal{D} = \{(t_i, \mathbf{y}_i)\}_{i=1}^s$. The data available at time $t_j \leq t < t_{j+1}$ are denoted by $\mathcal{D}_j = \{(t_i, \mathbf{y}_i))\}_{i=1}^j$.

2.2 APPROXIMATE BAYESIAN ESTIMATION

Bayesian state estimation reconstructs the posterior Probability Density Function (PDF) of the state, based on available data $\mathcal{D}_j$ and prior information $p(\mathbf{x}; t_0)$. For sequential measurements, this can be performed dynamically by using recursive Bayesian methods. An additional benefit is that, for every update of the belief state, it is not necessary to store the complete data set. In fact, all current information at time $t_j \leq t < t_{j+1}$ is embodied by the posterior $p(\mathbf{x}; t|\mathcal{D}_j)$.

Recursive Bayesian estimation consists of essentially two stages: prediction and update. The former uses the process model to predict the state PDF between one measurement time point t_j and the next. Given a continuous-time process model, the current belief state $p(\mathbf{x}; t|\mathcal{D}_j)$ at time $t_j \leq t < t_{j+1}$ is the solution of the Fokker-Planck equation for the extended process model, that is

$$\frac{\partial p}{\partial t} = -\nabla \cdot (fp) + \frac{1}{2}\Delta\left(\sigma p\right), \qquad (4)$$

where f governs the drift and σ is the extended diffusion tensor. The given initial condition for (4) is the latest posterior at time t_j, that is $p(\mathbf{x}; t_j|\mathcal{D}_j)$.

The extra information coming at time t_{j+1} reduces the uncertainty in the posterior, which is updated according to Bayes theorem. When the measurement $\mathbf{y}_{j+1}$ becomes available, the update step follows

$$p(\mathbf{x}; t_{j+1}|\mathcal{D}_{j+1}) = \frac{p(\mathbf{y}_{j+1}|\mathbf{x}; [t_0, t_j])p(\mathbf{x}; t_{j+1}|\mathcal{D}_j)}{p(\mathbf{y}_{j+1}|\mathcal{D}_j)}. \qquad (5)$$

This recurrence relation, together with (4), constitutes the basis for the optimal Bayesian solution. In general, this solution cannot be computed analytically and must be approximated, usually with Monte Carlo approaches. In the case of non-linear non-Gaussian recursive Bayesian estimation, sequential importance sampling/resampling techniques are commonly used (Arulampalam 2002).

The required posterior PDF can be approximated by a finite set of random samples $\{\mathbf{x}^{(i)}\}_{i=1}^{n_s}$ with associated

weights $\{w^{(i)}\}_{i=1}^{n_s}$ normalized to one. This approximation becomes the solution of the optimal Bayesian estimation as the number of samples tends to infinity. The trajectory $\mathbf{x}(t)$ of the process model for $t \in [t_0, t_{i+j}]$ can be approximated by the sampling $\mathbf{x}_{0:j+1}$.

Thus, the posterior can be estimated as

$$p(\mathbf{x}_{0:j+1}|\mathcal{D}_{j+1}) \approx \sum_{i+1}^{n_p} w_{j+1}^{(i)} \delta(\mathbf{x}_{0:j+1} - \mathbf{x}_{0:j+1}^{(i)}), \quad (6)$$

where $\{(\mathbf{x}_t^{(i)}, w^{(i)}(t))\}_{i=1}^{n_p}$ denotes a random measure that characterizes the posterior and δ is a Dirac delta. Given a proposal density $q(\mathbf{x}_{0:j+1}|\mathcal{D}_{j+1})$, it is possible to obtain samples $\mathbf{x}_{0:j+1}^{(i)}$ whose weights are

$$w_{j+1}^{(i)} \propto \frac{p(\mathbf{x}_{0:j+1}|\mathcal{D}_{j+1})}{q(\mathbf{x}_{0:j+1}|\mathcal{D}_{j+1})}. \quad (7)$$

If the importance proposal is chosen such that it can be factorized as

$$q(\mathbf{x}_{0:j+1}|\mathcal{D}_{j+1}) = q(\mathbf{x}_{j+1}|\mathbf{x}_{0:j}, \mathcal{D}_{j+1}) q(\mathbf{x}_{0:j}|\mathcal{D}_j), \quad (8)$$

then the existing samples can be iteratively augmented. This is be done by the following recursion

$$w_{j+1}^{(i)} \propto w_j^{(i)} \frac{p(\mathcal{D}_{j+1}|\mathbf{x}_k^{(i)}) p(\mathbf{x}_{j+1}^{(i)}|\mathbf{x}_j^{(i)})}{q(\mathbf{x}_{j+1}^{(i)}|\mathbf{x}_{0:j}^{(i)}, \mathcal{D}_{j+1})}. \quad (9)$$

By numerically integrating with the Euler-Maruyama method the SDE of the extended process model, we can sample from $p(\mathbf{x}_{j+1}^{(i)}|\mathbf{x}_j^{(i)})$. This distribution can be used as the importance proposal for the approximation, simplifying the recursive computation of the weights in (9) to

$$w_{j+1}^{(i)} \propto w_j^{(i)} p(\mathbf{y}_j|\mathbf{x}_{j+1}^{(i)}). \quad (10)$$

Since the form of $p(\mathbf{y}_j|\mathbf{x}_{j+1}^{(i)})$ is known from (3), the weights update requires only a simple function evaluation.

There are two main difficulties associated with sequential Monte Carlo approximations. First, the degeneracy problem is the common phenomenon where, after few iterations, all but few particles present negligible weight (Arulampalam 2002). A common measure of degeneracy is the Effective Sample Size (ESS), denoted by N_{eff} as defined as follows

$$N_{\text{eff}} = \frac{n_p}{1 + \text{var}[w_j^{*(i)}]} \approx \widehat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^{n_p} (w_j^{(i)})^2}. \quad (11)$$

The approximation is due to the fact that the quantity

$$w_j^{*(i)} = \frac{p(\mathbf{x}_j^{(i)}|\mathcal{D}_i)}{q(\mathbf{x}_{j-1}^{(i)}|\mathbf{y}_j)} \quad (12)$$

cannot be evaluated exactly. Small values of ESS are a symptom of degeneracy and it has been proved that the variance of the importance weights can only stochastically increase over time (Doucet 1998). The degeneracy can be easily detected, as shown in the experimental results, and can be solved by regenerating the sample set by resampling. However, standard resampling is the cause of the second difficulty with sequential importance sampling/resampling: the divergence caused by the collapse of multiple modes into a single degenerate configuration (King 2000). Our solution to that is a preventive clustering approach, which partitions the particles to contrast the predicted divergence. This is important not only for the estimation but also for the evaluation of any candidate intervention.

2.3 BELIEF STATE CLUSTERING

A dynamic clustering of the belief state will help us in three tasks: detecting multiple parameter configurations, avoiding the divergent behaviors of the importance resampling approximation, optimizing the quality of the designed intervention. We consider clusters in the posterior as alternative parameter configurations subject to local uncertainty. At any point in time, particles from the importance resampling approximation of the belief state are partitioned into n_c clusters, obtained with weighted K-means. This is done dynamically with cluster assignments recalculated at every Bayesian update. By initializing K-means with the previous centroids, the clusters can be efficiently tracked over time.

The problem of unimodal divergence is overcome by preventive partitioned resampling. More precisely, when the ESS falls below a certain threshold as a consequence of a Bayesian update step at time t_j, two Gaussians are fitted for every cluster. The first one, denoted as $\mathcal{N}^-$, locally approximates the belief state before t_j. Similarly, $\mathcal{N}^+$ is an approximation of the predicted belief state after the Bayesian update. After that, the preventive resampling is done separately for every cluster. This maintains the number of particles constant in every partition and avoids the degenerate loss of multimodality. The new particles are sampled from the proposal $\mathcal{N}^+$ and their new importance weights are given by $w^{(i)} \propto \mathcal{N}^-(\mathbf{x}^{(i)})/\mathcal{N}^+(\mathbf{x}^{(i)})$. The procedure is exemplified with 100 samples in Fig. 2. There, our strategy (right) outperforms the standard cluster resampling technique (left) in the approximation of the posterior probability, given a prior for parameter θ. This performance is due to the fact that, whereas standard techniques resample after the Bayesian update, we do that in advance. In fact, this allows us to concentrate the samples in regions where the contribution to the approximation is substantial, instead of dissipating them were the weights present

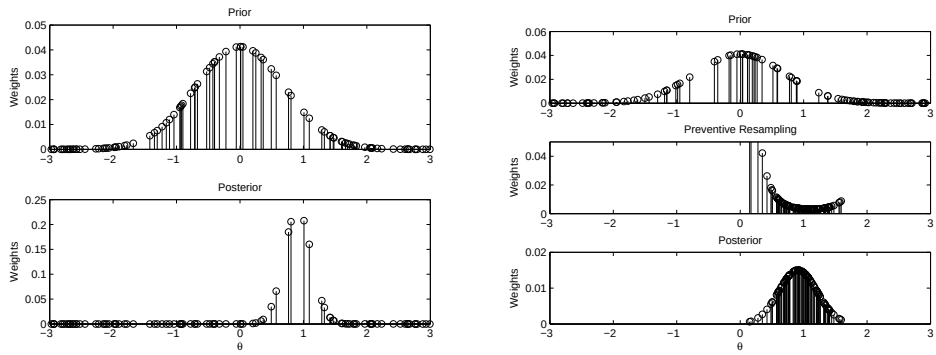

Figure 2: Compared with standard cluster resampling (left), our preventive approach better approximates the posterior PDF (right) for the parameter θ. This is due to an the intermediate step that provides a higher sample density the most significant regions of the parameter space.

diminished values.

In order to maintain in every cluster the number of samples constant over time, we forbid sample diffusion between adjacent clusters. This means that the new samples that previously belonged to cluster $\mathcal{C}_j$ will be rejected if their new resampled coordinate is in the partition assigned to cluster $\mathcal{C}_k$, for $k \neq j$. Convex regions of the parameter space are implicitly assigned according to a Voronoi partitioning based on the centroids of the clusters. The results show that this novel strategy significantly facilitates the convergence of the algorithm.

2.4 OPTIMIZED INTERVENTIONS

In order to define our design criterion, we need to quantify the number of alternative uncertain structures in our belief state. Assuming that our dynamical clustering acts upon the clusters of particles $\{\mathcal{C}_j\}_{j=1}^{n_c}$, we assign to every cluster its global weight. Let $\mathcal{P}_j$ be the partition associated with cluster $\mathcal{C}_j$, then the global weight $\mathcal{W}_j$ is given by

$$\mathcal{W}_j = \int_{\mathcal{P}_j} p(\mathbf{x}; t|\mathcal{D}) d\mathbf{x} \approx \sum_{i:\mathbf{x}^{(i)} \in \mathcal{P}_j} w^{(i)} p(\mathbf{x}^{(i)}; t|\mathcal{D}). \tag{13}$$

The vector $[\mathcal{W}_1, \ldots, \mathcal{W}_{n_c}]$ is normalized to unity since

$$\sum_{j=1}^{n_c} \mathcal{W}_j = \int p(\mathbf{x}; t|\mathcal{D}) d\mathbf{x} = 1 \tag{14}$$

and can be interpreted as a vector representation of a probability mass function p_c. At this point, its discrete entropy $H[p_c]$ can be easily computed plugging (13) into

$$H[p_c] = -\sum_{i=j}^{n_c} \mathcal{W}_j \log(\mathcal{W}_j). \tag{15}$$

By taking the following as information criterion to maximize $1/(1+H[p_c])$, we favor the belief states with the lowest $H[p_c]$, that correspond to the ones that maximally discriminate between alternative structures. Now, let $p(\mathbf{x}; t_j|\mathcal{D}_j)$ be the belief state at time t_j and let us consider an experimental design for the interval $[t_j, t_{j+l}]$. By using a certain u, for a given realization of the measurements $\mathbf{y}_{j:j+l}$, one can compute $H_u[p_c]$ following (15). We want to find the intervention u belonging to the function space $\mathcal{U}$ such that

$$u^* = \arg\max_{u \in \mathcal{U}} \ \mathbb{E}_{\mathbf{y}_{j:j+l}} \left[\frac{1}{1 + H_u[p_c]} \right] \tag{16}$$

subject to the following two constraints:

1. energy cost: $\mathrm{En}[u] < \alpha$, where En is an energy cost function and α is a given threshold;
2. non-decreasing KLD:

$$\phi(p_u) \geq \phi(p_0), \tag{17}$$

$\phi(p_u) = \mathbb{E}_{\mathbf{y}_{j:j+l}} [\mathrm{KL}(p_u(\mathbf{x}; t_{j+l}|\mathcal{D}_{j+l}))\|p(\mathbf{x}; t_j|\mathcal{D}_j)]$ and similarly for p_0. Here KL denotes the KLD function, while $p_u(\mathbf{x}; t_{j+l}|\mathcal{D}_{j+l})$ and $p_0(\mathbf{x}; t_{j+l}|\mathcal{D}_{j+l})$ denote the resulting belief state respectively with or without the intervention.

The former constraint imposes an upper bound on the energy cost for the intervention. The latter avoids the selection of degenerate interventions that stimulate unpredictable behaviors. In fact, these could finally hide the underlying dynamics in larger and more uncertain clusters, giving an illusory reduction of the number of alternative structures.

The estimation of the expected value in (16) and in the second constraint is computationally demanding or even infeasible. As shown in the results, it can be well approximated by taking the values which follow from the estimation of $\mathbb{E}[\mathbf{y}_{j:j+l}]$. The resulting optimization problem is generally non-convex and its

complexity depends significantly on the construction of the discrete representation of u and on its related dimensionality.

3 EXPERIMENTAL RESULTS

To illustrate the effectiveness of our approach, we applied it to a modified version of the Goodwin model. This is a biochemical negative feedback oscillator that describes the molecular mechanism of the circadian clock of Neurospora and Drosophila (Ruoff 2001). The process model, the measurement conditions and the nominal parameter values have been chosen for the sake of clarity: they present an interesting non-linear behavior and emerging multimodal posteriors. The multimodality persist with observations and clearly represent alternative kinetic structures. The most interesting non-linear features are the presence of a stabilizing feedback loop, of an inhibitory threshold effect and of two competing parallel reactions.

3.1 BIOCHEMICAL MODEL

The model is schematically represented in Fig. 3 and its mathematical formulation is the following:

$$d\begin{bmatrix}c_1\\c_2\\c_3\\c_4\end{bmatrix} = \left(\mathbf{N}\begin{bmatrix}r_1\\r_2\\r_3\\r_4\end{bmatrix} + \mathbf{G}\begin{bmatrix}u_1\\u_2\\u_3\\u_4\end{bmatrix}\right)dt + \mathbf{D}^{1/2}dW_t \quad (18)$$

where every c_i represent the concentration of the chemical species C_i, u_j are the components of the time-dependent input intervention and r_k are the following reactions

$$\begin{aligned} r_1(c_1,\theta_1) &= \theta_1 c_1 \\ r_2(c_1,\theta_2) &= \theta_2 c_1 \\ r_3(c_2,c_3,\theta_3) &= \theta_3 c_2 c_3 \\ r_4(c_4,\theta_4,\theta_5) &= \theta_4 c_4^2/(c_4^2+\theta_5^2). \end{aligned} \quad (19)$$

Reactions rates, concentrations and parameters are

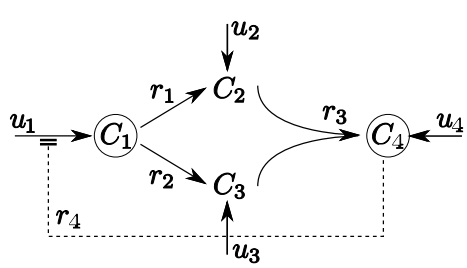

Figure 3: The modified Goodwin model.

here assumed to remain non-negative. The matrices $\mathbf{N}$, $\mathbf{G}$ and $\mathbf{D}$ respectively represent the stoichiometric matrix that governs the deterministic component of the reactions, the input interface matrix that maps the inputs into the states and, finally, the covariance of the Wiener process. These are defined as follows:

$$\mathbf{N} = \begin{bmatrix}-1 & -1 & 0 & -1\\ 1 & 0 & -1 & 0\\ 0 & 1 & -1 & 0\\ 0 & 0 & 6 & 0\end{bmatrix},\ \mathbf{G} = \begin{bmatrix}1\\1\\0\\1\end{bmatrix}, \quad (20)$$

while $\mathbf{D} = \mathrm{diag}[0.1, 0.05, 0.05, 0.05]$. The stoichiometric matrix $\mathbf{N}$ represents the net effect of all the involved reactions visualized in Fig. 3. Note the parallel structure of the competing reactions r_1, r_2: their parametrization is unknown and their products cannot be measured, creating alternative candidate structures. The matrix $\mathbf{G}$ interfaces the external interventions with the state space of the process model, describing the feasible manipulations.

The extended process model is obtained as described in Subsection 2.1. It is related to the Horn-Jackson-Feinberg formalism (Bullinger 2008) and its deterministic component is as follows:

$$\frac{d}{dt}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\\x_6\\x_7\\x_8\\x_9\end{bmatrix} = \begin{bmatrix} u_1 - x_6 - x_7 - x_9 \\ u_2 + x_6 - x_8 \\ x_7 - x_8 \\ u_4 + 6x_8 \\ 2x_4(u_4+6x_8) \\ r_1\frac{(u_1-x_6-x_7-x_9)}{x_1} \\ r_2\frac{(u_1-x_6-x_7-x_9)}{x_1} \\ r_3(\frac{(u_2+x_6-x_8)}{x_2} + \frac{(x_7-x_8)}{x_3}) \\ r_4(2\frac{(u_4+6x_8)}{x_4} - \frac{x_4}{x_5}) \end{bmatrix} \quad (21)$$

The measurement process is formalized as follows:

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}(t_i) + \mathbf{v}_i, \quad (22)$$

where the measurement matrix is

$$\mathbf{H} = \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\end{bmatrix} \quad (23)$$

and $\mathbf{v}_i$ is an additive white Gaussian noise process with covariance matrix $\mathrm{diag}[0.6, 0.6]$. This corresponds to the case of a process with reduced measurability, where only the initial and terminal chemical species can be directly observed by experimentation. Measurable species are represented within a circle in Fig. 3.

3.2 CLUSTER STABILITY

Our preventive cluster resampling allows us to avoid unimodal divergences, which are caused by recursive sample impoverishment in regions of the multimodal belief state. The results plotted in Fig. 4 show the dynamics of the belief state for the estimation of the score function. Standard resampling rapidly diverges (left), bringing more than 90% of the 500 realizations into degenerate absorbing states. Viceversa, our strategy successfully maintains more than 90% of the realizations

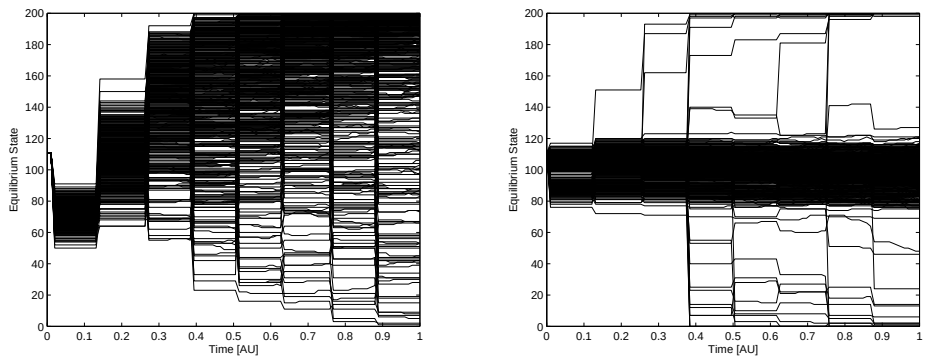

Figure 4: Convergence is achieved when the equilibrium state remains in the central region. Standard resampling exhibits frequent unimodal degeneration (left), while our clustered resampling scheme shows greater stability (right). Time is measured in arbitrary units.

in the central convergent region. This is a fundamental requisite for the approximation of the experimental criterion: without this stability, the utility value is systematically overestimated and, consequently, the design automation provides degenerate solutions.

3.3 STRUCTURE IDENTIFICATION

In our example, we consider the following parameter vector $\theta = [7, 1, 10, 3, 1]^{\mathrm{T}}$ and initial conditions $z_0 = [3, 0.6, 0.4, 0.8]^{\mathrm{T}}$. We dynamically perform the design, considering the interval between one sample and the next. Our approach is insensitive to the distribution of measurement time points, as long as they are sequential. In Fig. 6, we consider 10 incomplete measurements that are equally spaced for visualization purposes. For the sake of clarity, we activate our method at time $t = 0.5$, allowing a comparative evaluation of its advantages. The dynamics of the belief state for $\theta_1 - \theta_2$ is plotted in Fig. 6, on the left without intervention and on the right with our optimized intervention. The identification of the correct structure is rapid, almost creating a discontinuity in the evolution of the belief state. We considered the intervention space of independently scaled four dimensional Heaviside functions, shifted to the current point in time. The numerical optimization has been performed with simulated annealing, constrained by the following energy bound

$$\mathrm{En}[u] = \|u\|_2 < 20. \tag{24}$$

The trajectory of the dynamical system subject to the optimal intervention is plotted in Fig. 5.

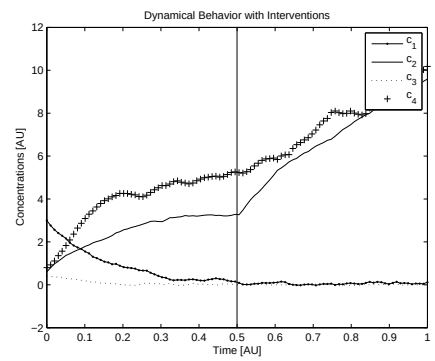

Figure 5: Trajectory of the dynamical system subject to the optimal intervention. Time and concentrations are measured in arbitrary units.

4 DISCUSSION

In this study we proposed and evaluated a new methodology for structure identification by designing optimized interventions for a given experiment. The new approach enables the selection between alternative explanations and provides correct results in cases where standard experimental design fails. It is especially suited for problems where multiple parameter configurations persists, for example due to intrinsic properties of the system or because of data scarcity. Our strategy is based on a novel preventive resampling technique, that provides stability to the simulated inference and feasible approximations for the design criterion to optimize. The sample clusters represent alternative structures that are candidate explanations for the measured system dynamics. They are used to estimate the intervention that maximizes the expected difference between the probabilities associated with the structures. This is done by considering two additional constraints for the interventions: an upper bound on the energy cost and the restriction to reasonably predictable behaviors.

The three main advantages over the existing methods

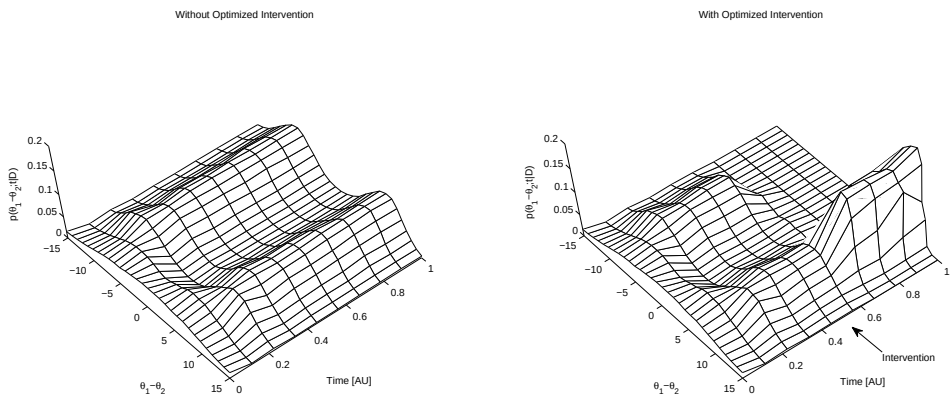

Figure 6: Our design approach permit the disambiguation between alternative structures (right) in cases where multimodality would otherwise persist (left). After its activation at time $t = 0.5$ it identifies the correct mode, excluding the alternative candidate explanation.

are the following. First, the stability of the approximated utility function is due to the clustering of the belief state. Second, the ability to constrain the interventions avoids unpredictable trajectories. Third, it combines precision in the parameter estimation and simultaneous selection between alternative structures. These capabilities will prove useful especially in cases where the measurement process presents reduced observability and complex dynamics. The resulting optimization problem can be computationally demanding but we demonstrated that its numerical solution is feasible for realistic applications. An interesting observed phenomenon is the rapidity of the discrimination between alternative structures. This shows that promising results can be obtained with almost reactive, short term design.

While we show that our new methodology is accurate and useful, we note that it is sensitive to the selection of the intervention space. There are couple of directions for future research. First, the assignment of problem-specific intervention spaces and energy costs could significantly simplify the optimization process. Second, the determination of the number of sample clusters is a task that deserves attention due to its central role in the design criterion. These are both the subjects of ongoing research.

Acknowledgments

We thank Jörg Stelling and Cheng Soon Ong for insightful discussions and the anonymous reviewers for helpful comments. This project was financed with a grant from the Swiss SystemsX.ch initiative, evaluated by the Swiss National Science Foundation.

References

S. Akaho and K. Fukumizu (2007). Active Learning for Network Estimation. *Proc. of the 2007 IEEE Symp. on Computational Intelligence in Bioinformatics and Computational Biology*: 402–409.

M.S. Arulampalam, S. Maskell, N. Gordon and T. Clapp (2002). A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Trans. on Signal Processing* **2**(50): 174–188.

E. Bullinger, D. Fey, M. Farina and R. Findeisen (2008). Identifikation Biochemischer Reaktionsnetzwerke: ein Beobachterbasierter Ansatz. *Automatisierungstechnik* **5**(56): 169–279.

A. Doucet (1998). On Sequential Simulation-Based Methods for Bayesian Filtering. Technical Report, Department of Engineering, University of Cambridge.

A. Doucet and V. Tadić (2003). Parameter Estimation in General State-Space Models using Particle Methods. *Annals of the Institute of Statistical Mathematics* **2**(55): 409–422.

O.D. King and D.A. Forsyth (2000). How Does CONDENSATION Behave with a Finite Number of Samples? *Proc. of the European Conf. on Computer Vision*: 695–709.

P. Ruoff, M. Vinsjevik, C. Monnerjahn and L. Rensing (2001). The Goodwin Model: Simulating the Effect of Light Pulses on the Circadian Sporulation Rhythm of Neurospora crassa. *Journal of Theoretical Biology* **1**(209): 29–42.

M. Seeger, F. Steinke and K. Tsuda (2007). Bayesian Inference and Optimal Design in the Sparse Linear Model. *Proc. of the 11th Int. Conf. on Artificial Intelligence and Statistics*: 444–451.

F.X. Wu, L. Mu and R. Luo (2008). Complexity Analysis and Optimal Experimental Design for Parameter Estimation of Biological Systems. *Canadian Conf. on Electrical and Computer Engineering*: 393–397.