

Reduction and identification of dynamic models. Simple example: generic receptor model.

Heikki Haario,[†] Leonid Kalachev,^{*} Marko Laine,^{††}

[†] Lappeenranta University of Technology, Lappeenranta, Finland

^{*} University of Montana, Missoula, MT, USA

^{††} Finnish Meteorological Institute, Helsinki, Finland

Abstract

We consider a general scheme for reduction and identification of dynamic models using available experimental data. Analysis of reliability regions for estimated parameter values is performed using Markov Chain Monte Carlo simulation methods. In cases where some of the model parameters are not reliably defined, and when the values of certain model parameters turn out to be small (or large), asymptotic reduction techniques are used to reduce the models (i.e., to reduce the number of equations, number of reliably identifiable parameter, etc.). Consecutive application of parameters estimation (together with their reliability regions) and asymptotic reduction procedures will produce the new simpler model with the smallest number of parameters reliably identifiable by the available data (i.e., the model that is optimal with respect to available data). The ideas are illustrated using a simple example related to bio-medical applications: a model of a generic receptor.

Keywords: Model identification, asymptotic methods, Boundary Function Method, model reduction, Markov chain Monte Carlo (MCMC)

1 Introduction

One of the important goals of contemporary mathematical biology is to produce and analyze the models that provide quantitative description of the behavior of biological systems based on real experimental data. Identification of such models, i.e., statistically reliable estimation of model parameters that would lead to statistically reliable model predictions of future events for a given biological system, is often complicated by the fact that the available experimental data from field measurements is noisy or incomplete. Moreover, models typically are nonlinear, they may be complex, and contain a large number of correlated parameters. As a result, the parameters are poorly identified by the data, and the reliability of the model predictions may be questionable. In this paper we want to point out what the researchers working in biological, medical, etc., fields should look for to identify such problematic situations, and how to overcome these problems.

To evaluate the accuracy of the estimated parameters in a non-linear multiparameter model, it is important to consider possible cross-correlation and identifiability of the parameters. In classical statistical analysis the error estimates are approximative, based on a linearization of the model, and may sometimes be quite misleading. Moreover, the question of the reliability of the model predictions, i.e., how the uncertainty in model parameters is reflected in the model response, is often left open. Both of these problems may be properly treated by Markov chain Monte Carlo (MCMC) methods. Using MCMC methods, the estimation of model parameters and predictions are performed according to the Bayesian paradigm. All uncertainties in the data as well as the modelling results are treated as random variables that have statistical distributions. Instead of a single fit to the data, 'all' parameter combinations of the model that, statistically, fit the data 'equally well' are determined. A distribution of the unknown parameters is generated using available prior information (e.g. previous studies) and statistical knowledge of the observation noise. Computationally, the distribution is generated using the Markov chain Monte Carlo (MCMC) sampling approach. Up-to-date adaptive computational schemes may be employed in order to make the simulations as effective as possible.

A model must fit available experimental data reasonably well. For an improved fit another, often more complex, model containing more parameters is tried. For reliable simulations, the uncertainty of the parameter values must be small enough. If this is not the case, some new experiments may be designed based on model analysis that would indicate which additional experimental data must be collected for identification of parameters. For methods on optimal design of experiments see, e.g., [1]. In many practical situations, however, sufficient data for accurate estimation of all the model parameters is simply not available. This situation is the focus of the present paper. We address the question: how to arrive at a reduced model (with properly identified model parameters), that captures the key features of the studied phenomena well enough.

To find out how well complex models are actually supported by available data, one must be able to both properly analyze the identifiability of the parameters of a given model, and, if needed, to modify the model itself. For this purpose, we present a process for model selection that consists of two main steps, MCMC sampling and model reduction. To clearly point out the methodology, we use a simple generic receptor model with synthetic data as our example. For a more complex case with real data, see [8].

When parameters and their distributions are estimated, some of the (non-dimensionalized) parameters may turn out to be small, while others may be large; reliability regions for some parameter values may be tight and convex, while for others they may have of a shape of a line segment or a "banana" shape (which would indicate that some parameters are correlated or not well defined), etc. The idea then is to eliminate the parameters with the smallest impact on the model behavior. In a straightforward case, a parameter may simply be set to zero. In more complex situations, e.g., in case of large but unidentified parameters, or parameters which may be grouped together, we employ a structural reduction of the model using an asymptotic approach.

The reduced model produced by application of some asymptotic procedure (e.g., Boundary Function Method [5], Matching techniques [6], [7], etc.) should then be fitted to available data again. The estimates of new (smaller) parameter set should be

statistically more reliable. Consecutive application of parameters' and their reliability regions' estimation and asymptotic reduction is expected to eventually produce simpler model with the smallest number of parameters reliably identifiable by the data (i.e., the model that is optimal with respect to available data).

There are several important (and widely known) features of the dynamic models that are related to bio-medical, chemical kinetics, and other applications that we would want to emphasize before presenting the statement of the problem and its analysis:

(a) Many processes in biological and chemical kinetics are readily described using, the, so-called, kinetic *reaction schemes*. The kinetic schemes indicate which species interact with each other, which species and in what proportions appear or disappear during the reactions, etc. It is important to emphasize that the form of a kinetic reaction scheme does not completely define the dynamics of the model system: different choices of parameter values may lead to completely different types of model solution behavior. The qualitative types of solution behavior may change when some parameters change beyond certain, so-called, bifurcation values. As a result, the number of the steady states of the dynamic system, the stability properties of steady states, etc., also change. In what follows, we describe the methodology which is applicable to the models whose estimated parameter values stay away from their possible bifurcation values. Identification of models near bifurcations is a separate topic and it must be addressed separately.

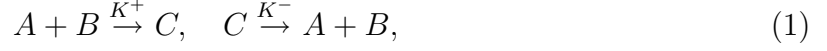
(b) The shapes of observed model solution curves (even when parameter estimates are away from their bifurcation values) depend on the characteristic time interval of interest. The duration of characteristic time interval is determined, e.g., by the time during which experimental data is collected. Often the choice of characteristic time interval determines which model parameters (related to characteristic times of processes that occur in a biological system) can or cannot be estimated reliably.

The paper is organized as follows. In Section 2 we present the model describing generic receptors. In Section 3 we fit the model to three data sets corresponding to cases of moderate, fast and slow forward and reverse binding reactions. We also use Markov Chain Monte Carlo (MCMC) to produce probability distributions for estimated parameter values with the goal of better understanding possible indications that the original model may not be reliably identifiable from the available data. More details on both MCMC algorithms and asymptotic model reduction methods are presented in the Appendices.

2 Generic receptors: model formulation

Generic receptor model is widely used in pharmacokinetics (e.g., to describe protein binding during drug propagation / distribution), in neuroscience (to approximately describe binding of glutamate and other neurotransmitters to real, more complex, receptors), in description of certain catalytic reactions in chemical engineering, etc. While the actual receptors in many applications may have very complex structure and are characterized by complicated reaction kinetic schemes, the generic receptors (often used for approximating the action of real receptors) are characterized only by two simplest reaction steps, forward and reverse binding of a given species to a receptor. Corresponding kinetic reaction

scheme for a generic receptor has the form:



where A denotes one molecule of certain species, B corresponds to a free receptor, and C represents a bound receptor; K^+ and K^- are the rate constants of forward and reverse binding reaction.

Let us introduce notations $[A]$, $[B]$, and $[C]$ for concentrations of A , B , and C . Then, applying the Law of Mass Action [?] we can immediately write down the system of differential equations for concentrations that corresponds to (1):

$$\begin{aligned} \frac{d[A]}{d\tilde{t}} &= -K^+ \cdot [A] \cdot [B] + K^- \cdot [C], \\ \frac{d[B]}{d\tilde{t}} &= -K^+ \cdot [A] \cdot [B] + K^- \cdot [C], \\ \frac{d[C]}{d\tilde{t}} &= +K^+ \cdot [A] \cdot [B] - K^- \cdot [C], \\ 0 &\leq \tilde{t} \leq T. \end{aligned} \quad (2)$$

$$(3)$$

Here \tilde{t} is the time variable defined on some characteristic time interval $[0, T]$, where the value of T is determined by the context of the problem. For example, this could be the time interval during which the experimental data is collected. We supply (2) with the initial conditions (without loss of generality we assume that the initial concentration of A is greater than that of B):

$$[A](0) = A^* > 0, \quad [B](0) = B^* < A^*, \quad [C](0) = 0. \quad (4)$$

Let us non-dimensionalize the problem (2) – (4) by introducing the following new variables and parameters:

$$u = \frac{[A]}{A^*}, \quad v = \frac{[B]}{A^*}, \quad w = \frac{[C]}{A^*}, \quad (5)$$

$$t = \frac{\tilde{t}}{T}, \quad v^* = \frac{B^*}{A^*}, \quad k^+ = K^+ \cdot T \cdot A^*, \quad k^- = K^- \cdot T. \quad (6)$$

Using (5), (6) we re-write (2) – (4) in the form:

$$\begin{aligned} \frac{du}{dt} &= -k^+ \cdot u \cdot v + k^- \cdot w, \\ \frac{dv}{dt} &= -k^+ \cdot u \cdot v + k^- \cdot w, \\ \frac{dw}{dt} &= +k^+ \cdot u \cdot v - k^- \cdot w, \end{aligned} \quad (7)$$

$$0 \leq t \leq 1; \quad (8)$$

$$u(0) = 1, \quad v(0) = v^* < 1, \quad w(0) = 0. \quad (9)$$

Let us note that the duration of the non-dimensional time interval of interest (8) now is 1. Characteristic times associated with the forward and reverse binding reactions are proportional to $1/k^+$ and $1/k^-$. So, when we say that, e.g., forward binding is fast, this means that $1/k^+ \ll 1$ (or $k^+ \gg 1$), or if we say that the reverse binding is slow, this means that $1/k^- \gg 1$ (or $k^- \ll 1$), etc.

The problem (7) – (9) can be solved explicitly. However, for illustrative purposes, below let us consider only numerical solutions of this problem. To elucidate the structure of the dynamical system, let us also present its simplification related to substitution of the original system of three differential equations by the equivalent one differential equation for u and two algebraic (conservation) relations that allow us to find v and w when u is known.

We note that subtracting the second equation from the first, and adding the first and the third equations in (7), we obtain:

$$\frac{du}{dt} - \frac{dv}{dt} = 0, \quad \frac{du}{dt} + \frac{dw}{dt} = 0. \quad (10)$$

Integrating (10) with initial conditions (9), we obtain

$$v(t) = u(t) - 1 + v^*, \quad (11)$$

$$w(t) = 1 - u(t). \quad (12)$$

Substituting (11), (12) in (7), we finally arrive at the problem for u :

$$\frac{du}{dt} = -k^+ \cdot u \cdot (u - 1 + v^*) + k^- \cdot (1 - u), \quad 0 \leq t \leq 1; \quad (13)$$

$$u(0) = 1. \quad (14)$$

The models (7) – (9) and (11) – (14) are equivalent in the sense that they can be used interchangeably for data fitting and analysis of the results.

The resulting non-dimensionalized model contains three parameters, k^+ , k^- , and v^* . Assume that we have the knowledge that the original model provides realistic description of the system that we study, and that these parameters are not originally known and must be estimated from the data. Let us discuss what types of experimental data may correspond to (7) – (9) (or to its equivalent representation (11) – (14)), and how different experimental data will affect the problem of parameter identification.

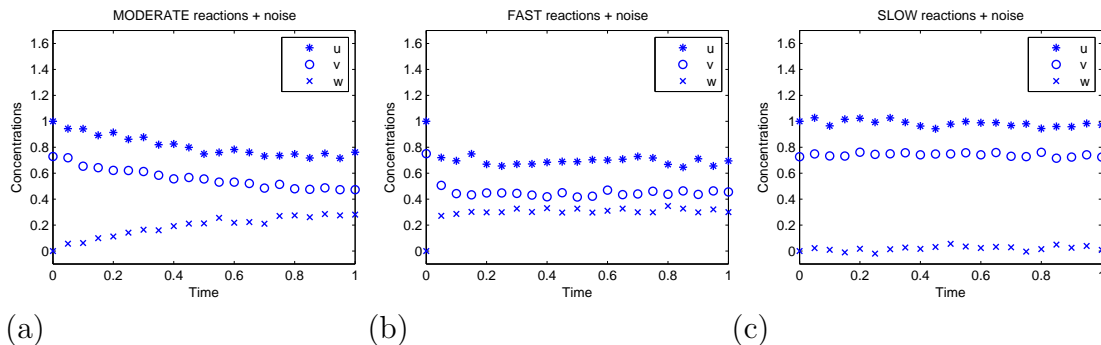


Figure 1: These data sets produced using the model (7) – (9) under various conditions. We will refer to them as the data sets corresponding to cases of (a) moderate reactions, (b) fast reactions, (c) slow reactions.

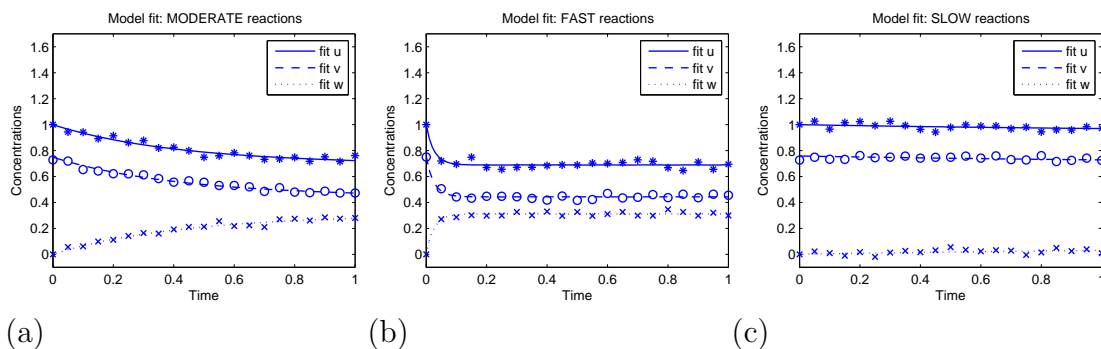


Figure 2: Model solutions fitted to data sets corresponding to cases of (a) moderate reactions, (b) fast reactions, (c) slow reactions.

3 Types of experimental data, model identification and model reduction

The same generic receptor reaction kinetic scheme (1) that produces the dynamic model (7) – (9) for particular choices of parameter values may generate, e.g., the data sets shown in Figures 1(a), 1(b) and 1(c). The data is collected on a time interval scaled to $[0, 1]$ at the time points uniformly distributed over this interval with the time step $\Delta t = 0.05$. The data sets are produced by numerically solving (7) – (9) with some fixed values of k^+ , k^- , and v^* (that are assumed to be unknown for the purpose of model identification) and adding the Gaussian noise with mean 0 and standard deviation $\sigma = 0.02$ to these numerical solutions. It is assumed that the measurements for u , v , and w are taken independently (so that their errors are not related, e.g., via expressions (11), (12)).

1. The model fit to data shown in Figure 1(a) is presented in Figure 2(a). It is performed using a standard optimization code (MATLAB solver *fitnonlin.m* or *fminsearch*, the same optimization codes are used for model fits mentioned below). Estimated parameter values are $k^+ = 0.9421$, $k^- = 0.9164$, $v^* = 0.7423$. Due to the fact that the estimated

rate constants are both moderate (they are of the same order as the duration of the time interval of interest, $[0, 1]$, and as the estimated value of v^*), we call this case the *moderate reactions model*, and we refer to corresponding data set as the *moderate reactions data set*. The estimates of reliability regions for model parameters are obtained using MCMC sampling techniques ([2]). In particular, here a standard Metropolis algorithm is used, with a Gaussian proposal distribution with covariance matrix given by the standard model linearization at the fitted parameter values. The parameter chain of 5000 was generated. A sample of accepted parameter values from the chain as well as the projections of the boundaries of 50% and 95% confidence regions on the parameter planes are shown in Figure 3(a). We note the following indicators of reliable parameter identification: (i) The shapes of the confidence regions' boundaries constructed for chain projections onto the parameter planes are convex (ideally, ellipse shaped). (ii) The means of the chains (shown by circles in Figure 3(a)) are close to or coincide with the originally estimated parameter values (shown by stars in Figure 3(a)). (iii) The chain is well mixed, i.e., the parameter values from the chain plotted against the sample number within the chain have the appearance of *Gaussian noise* or *white noise* (see Figure 3(b)). (iv) Finally, the 95% confidence bands for means of future experiments and 95% prediction bands for the outcomes of future single experiments (obtained from the model using parameter values sampled from the chain and sampled errors with statistical properties related to those of available experimental data) must be sufficiently narrow (see Figure 4). For reliably identified model parameters these bands must stay sufficiently narrow also for the cases when the current model setup is slightly changed, e.g., if the originally fixed coefficients in the model, such as the initial conditions for u and w , are changed, or when the current model is used as a part of a larger more general model formulation with known coefficients. In the latter case the predictions of the behavior for the more general model must be done using the constructed above distributions (chain) of the original model parameter values.

Estimated standard deviations for parameters k^+ , k^- , and v^* are $\sigma_{k^+} = 0.063449$, $\sigma_{k^-} = 0.16589$, and $\sigma_{v^*} = 0.0064356$, respectively. The results indicate that the data shown in Figure 1(a) allows one to reliably identify all three model parameters (considered originally unknown). In this case no additional experiments are required for further model clarification.

2. The model fit to data shown in Figure 1(b) is presented in Figure 2(b). Originally estimated parameter values are $k^+ = 22.6152$, $k^- = 23.1844$, $v^* = 0.7478$. Since the estimated nondimensional rate constants are both large (compared to the duration of the time interval of interest, as well as to the estimated value of v^*), we call this case the *fast reactions model*, and we refer to corresponding data set as the *fast reactions data set*. Once again we use the standard Metropolis MCMC sampling algorithm (with covariance matrix constructed by the same method as that used in the case of *moderate reaction model*) for estimating the reliability regions for model parameter values.

Let us check the indicators of reliable parameter identification: (i) We can immediately see that the shape of the reliability regions' boundaries on (k^+, k^-) -plane (i.e., the estimated boundaries of the 50% and 95% confidence regions for chain projections on this parameter plane) implies that the two parameters are correlated; see Figure 5(a). The minimal values in chain generated for k^+ , k^- are $\min_{chain}(k^+) = 16.9605$,

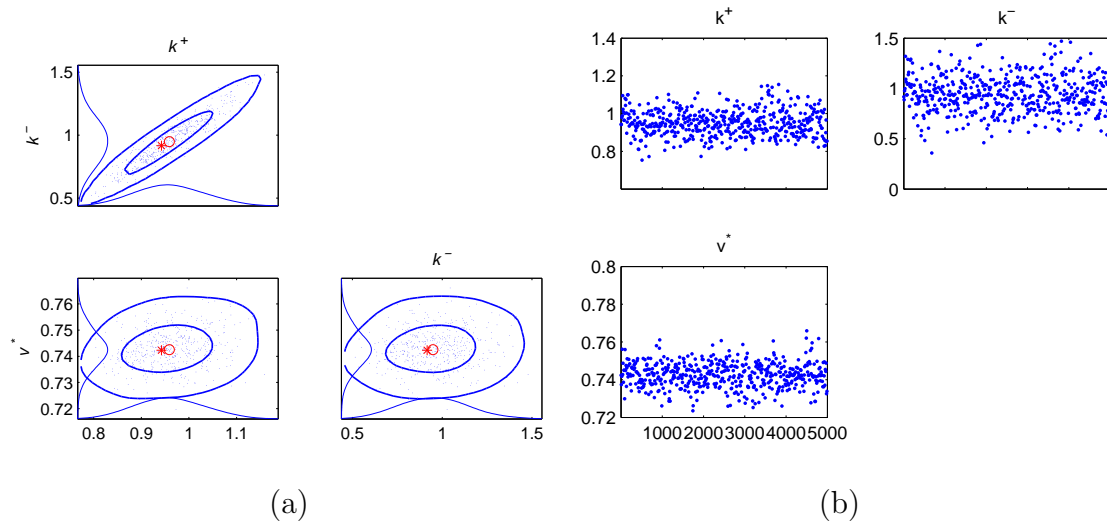


Figure 3: Case of moderate reactions. (a) The estimated boundaries of the 50% and 95% confidence regions for chain projections on parameter planes. Stars indicate the parameters values obtained by applying MATLAB optimization software; circles show the means of MCMC chains. (b) MCMC chains for model parameters.

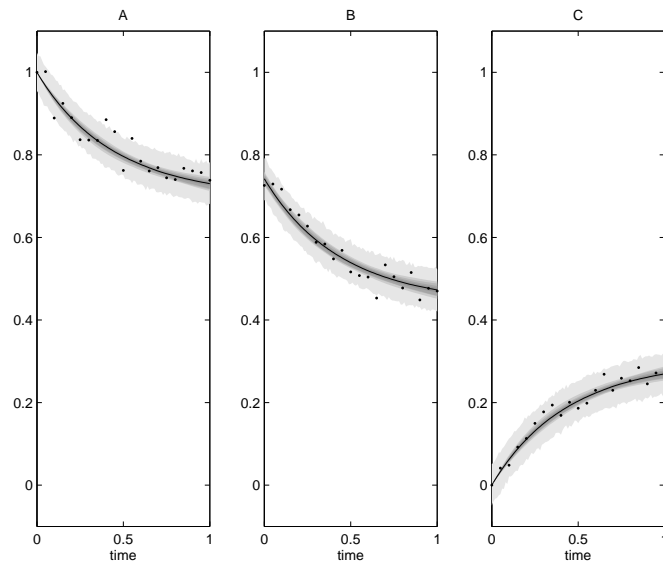


Figure 4: Case of moderate reactions. Dots correspond to data points; solid black curves approximate the medians of model solutions obtained for 500 parameter value sets sampled from the chain. Model predictions: 95% confidence bands for means (dark grey; practically undistinguishable from solid black curves) and 95% prediction bands for outcomes of single experiments (light grey).

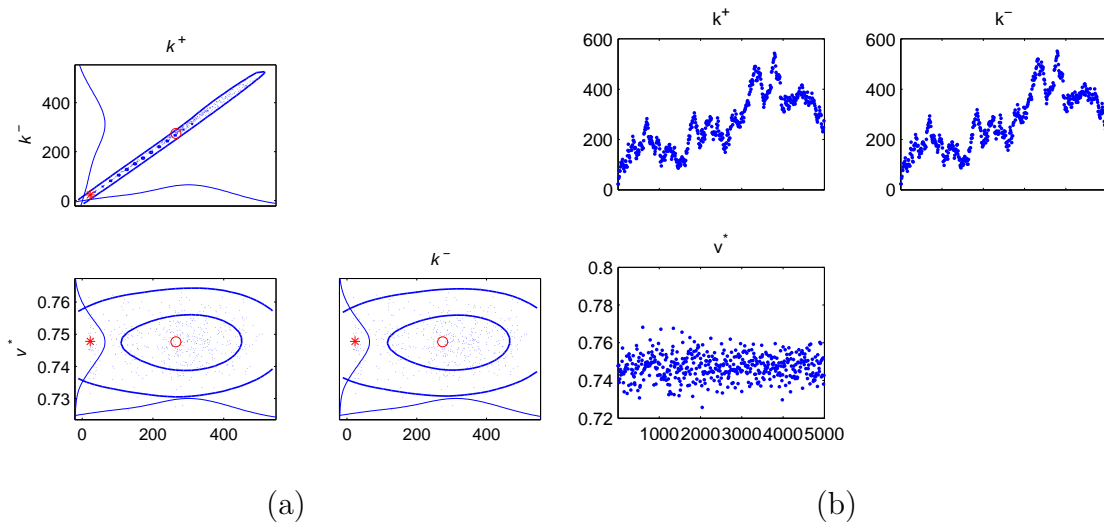


Figure 5: Case of fast reactions. (a) The estimated boundaries of the 50% and 95% confidence regions for chain projections on parameter planes. Stars indicate the parameters values obtained by applying MATLAB optimization software; circles show the means of MCMC chains. (b) MCMC chains for model parameters.

$\min_{chain}(k^-) = 16.9471$, respectively. Since these values are much greater compared to the estimated value of v^* , we will be able to use an asymptotic reduction procedure to produce a new model that contains only two parameters that are not correlated; see discussion below. (ii) The means of the chains (shown by circles in Figure 5(a)) are not at all close to the originally estimated parameter values (shown by stars in Figure 5(a)). (iii) The chain appears to be not well mixed, i.e., the parameter values from the chain plotted against the sample number within the chain do not have the appearance of *Gaussian noise* or *white noise* (for k^+ and k^- ; see Figure 5(b)); (iv) The 95% confidence bands for means of future experiments and 95% prediction bands for the outcomes of future single experiments (obtained from the model using parameter values sampled from the chain as well as sampled errors with statistical properties related to those of available experimental data) are sufficiently narrow (see Figure 6(a)) for the major part of the time interval of interest. However, if we are interested in model predictions for short time intervals, the uncertainty of predictions on such time intervals becomes large (see Figure 6(b)).

Evidently, the available data does not allow one to reliably identify all the three parameters. To eliminate uncertainty in estimation of k^+ and k^- additional experimental observations on a shorter time interval must be taken (e.g., taking experimental measurements at a number of uniformly spaced points on the interval $[0, 0.05]$ may be suggested). If no additional measurements are done, an equivalent (with respect to the available data) simpler model containing fewer reliably identifiable parameters may be constructed.

To reduce the original model, and check how many (and which) parameters are iden-

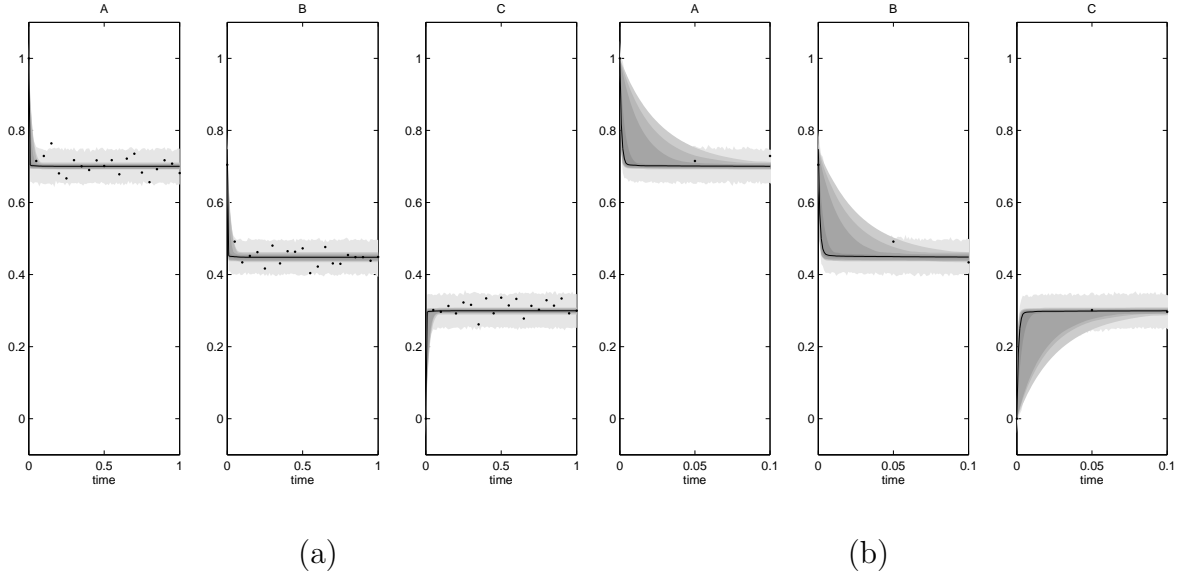


Figure 6: Case of fast reactions. Dots correspond to data points; solid black curves approximate the medians of model solutions obtained for 500 parameter value sets sampled from the chain. Model predictions: 95% confidence bands for means (dark grey) and 95% prediction bands for outcomes of single experiments (light grey). (a) Original time interval. (b) Short time interval.

tifiable, we note that estimated values of rate constants are large compared to estimated moderate value of v^* : $k^+ \gg v^*$, $k^- \gg v^*$. This means that the characteristic times associated with the forward and reverse binding reactions are short *compared to the characteristic time interval*: $1/k^+ \ll 1$, $1/k^- \ll 1$. Let us introduce a small parameter $0 < \varepsilon \ll 1$ so that the new re-scaled non-dimensional rate constants \tilde{k}^+ , \tilde{k}^- become moderate:

$$\tilde{k}^+ = \varepsilon \cdot k^+ = O(1), \quad \tilde{k}^- = \varepsilon \cdot k^- = O(1). \quad (15)$$

Notation $O(1)$ means that as $\varepsilon \rightarrow 0$, asymptotically we have $\lim_{\varepsilon \rightarrow 0} \tilde{k}^\pm = C$, where C is a constant independent of ε . Now the fact that some parameters are small (or large) in the original non-dimensional model will be shown explicitly using the small parameter factor. After substitution of (15) into (11) – (14) the new re-scaled model can be written as follows:

$$\varepsilon \frac{du}{dt} = -\tilde{k}^+ \cdot u \cdot (u - 1 + v^*) + \tilde{k}^- \cdot (1 - u), \quad 0 \leq t \leq 1; \quad (16)$$

$$u(0) = 1; \quad (17)$$

with two additional relations (11), (12).

The above model formulation represents a classical example of a *singularly perturbed problem*. A brief discussion of the conditions under which a model formulated in terms of

a more general singularly perturbed system may be reduced are given in Appendix 1 (see also Vasil'eva et al. [5]). Here we just mention that the reduction of (11), (12), (16), and (17) on an open interval $0 < t \leq 1$ may be performed by just setting $\varepsilon = 0$. We end up with a new model formulation:

$$\bar{u} \cdot (\bar{u} - 1 + \bar{v}^*) + K \cdot (1 - \bar{u}), \quad \bar{v} = \bar{u} - 1 + v^*, \quad \bar{w} = 1 - \bar{u}; \quad 0 < t \leq 1. \quad (18)$$

Here $K = \tilde{k}^-/\tilde{k}^+$, and by virtue of (15), we also have that $K = k^-/k^+$. We use notation \bar{u} , \bar{v} , and \bar{w} to denote the approximations to u , v , and w in an open interval $0 < t \leq 1$.

We note that the new *reduced* model (18) contains not three but only two parameters: K and v^* . Let us re-fit the new model to available data. The two parameter values $K = 1.0336$, $v^* = 0.7478$ are now reliably defined: (i) The shapes of the confidence regions projections onto the parameter plane are convex (ideally, ellipse shaped); see Figure 7(a). (ii) The means of the chains (shown by circles in Figure 7(a)) are close to the originally estimated parameter values (shown by stars in Figure 7(a)). (iii) The chain is well mixed, i.e., the parameter values from the chain plotted against the sample number within the chain have the appearance of *Gaussian noise* or *white noise* (see Figure 7(b)). (iv) Model predictions, however, are still reliable only for the time intervals that do not include some vicinity of the initial instant of time, e.g., for $[0.05, 1]$. Since no new experimental data was used, the predictability properties of the new model on the short time interval $[0, 0.05]$ cannot be improved compared to the original model.

Remark. To estimate the parameters of the reduced model several approaches may be used. One of the approaches is related to direct estimation of two parameters in algebraic system (18). Yet another approach (that involves taking *numerical limit*) is related to fixing the value of $k^+ > \min_{chain}(k^+) = 16.9605$, and estimating the two parameters from the resulting model that contains differential equation (16), written as

$$\frac{du}{dt} = k^+ \cdot (-u \cdot (u - 1 + v^*) + K \cdot (1 - u)), \quad 0 \leq t \leq 1; \quad (19)$$

together with (11), (12), (17). We used the latter approach for parameter estimation.

3. The model fit to data shown in Figure 1(c) is presented in Figure 2(c). Originally estimated parameter values are $k^+ = 0.0326$, $k^- = 0.0458$, $v^* = 0.7555$. Since the estimated rate constants are both small (compared to the duration of the time interval of interest, as well as to the estimated value of v^*), we call this case the *slow reactions model*, and we refer to corresponding data set as the *slow reactions data set*. Same as before, we use the standard Metropolis MCMC sampling algorithm (with approximate covariance matrix constructed using model linearization) for estimating parameter values' reliability regions.

Indicators of reliable parameter identification mentioned in the previous two cases now show us the following: (i) The estimated boundaries of the 50% and 95% confidence regions for chain projection on (k^+, k^-) -plane indicate that k^+ is comparatively small (and thus, the model may be reduced by taking, e.g., the limit as k^+ tends to zero), and that k^- is not well defined (since k^- may also be small, another possible model reduction could

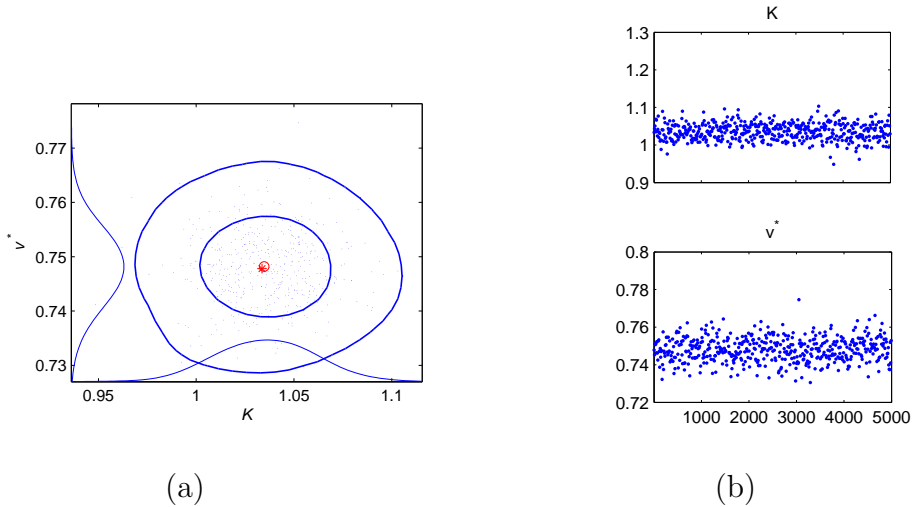


Figure 7: Case of fast reactions. (a) The estimated boundaries of the 50% and 95% confidence regions for chain projection on a parameter plane for reduced model that contains two parameters instead of original three. Stars indicate the parameters values obtained by applying MATLAB optimization software; circles show the means of MCMC chains. (b) MCMC chains for reduced model parameters.

be obtained by taking the limit as k^- approaches zero); see Figure 8(a). The minimal and maximal values in the chain generated for k^+ , k^- are: $\min_{chain}(k^+) = 0.0115 < k^+ < \max_{chain}(k^+) = 0.2547$, $\min_{chain}(k^-) = 0.0033 < k^- < \max_{chain}(k^-) = 8.5201$. Below we will use an asymptotic reduction procedure to arrive at two simpler models that cannot be distinguished by the available data. Only one parameter can be reliably identified without additional experiments. (ii) The means of the chains (shown by circles in Figure 8(a)) are not close to the originally estimated parameter values (shown by stars in Figure 8(a)). (iii) The chain appears to be not well mixed, i.e., the parameter values from the chain plotted against the sample number within the chain do not have the appearance of *Gaussian noise* or *white noise* (for k^+ and k^- ; see Figure 8(b)); (iv) The 95% confidence bands for means of future experiments and 95% prediction bands for the outcomes of future single experiments (obtained from the model using parameter values sampled from the chain as well as sampled errors with statistical properties related to those of available experimental data) are sufficiently narrow (see Figure 9(a)) for the time interval of interest. However, this only remains true if we are going to repeat absolutely the same experiment (using the system with the same initial conditions). If we slightly change the experimental setup (e.g., change the original initial conditions to other physiologically relevant ones, e.g., with $u(0) = 0.5$ and $w(0) = 0.5 \neq 0$), the model predictions made with the parameter distributions constructed earlier will not be reliable any more (see Figure 9(b) for uncertainties of predictions). For numerical computation of the slightly changed model predictions we used the original non-dimensionalized model formulation (7), (8) with the new initial conditions mentioned above.

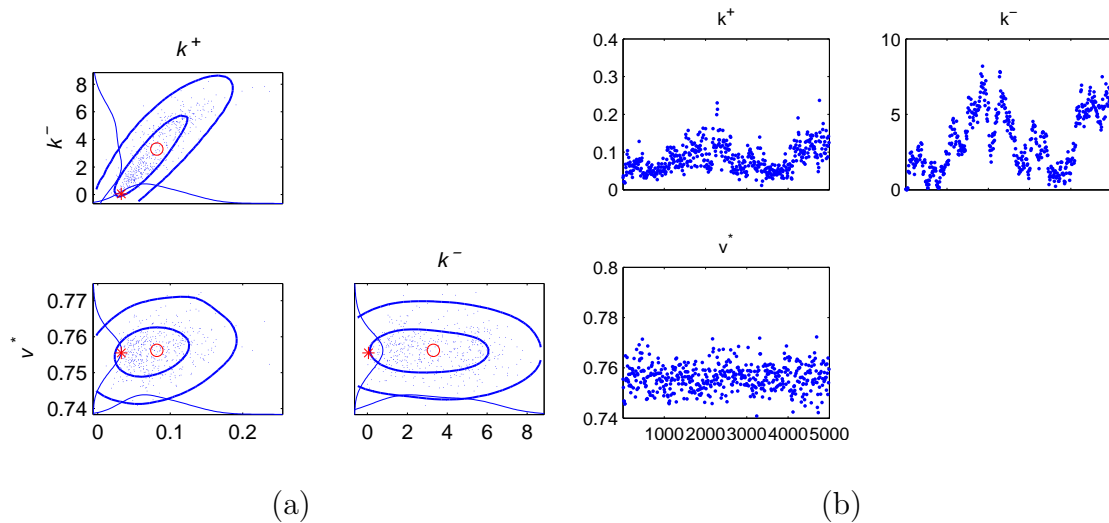


Figure 8: Case of slow reactions. (a) The estimated boundaries of the 50% and 95% confidence regions for chain projections on parameter planes. Stars indicate the parameters values obtained by applying MATLAB optimization software; circles show the means of MCMC chains. (b) MCMC chains for model parameters.

Thus, the available data does not allow one to reliably identify all the three parameters. To eliminate uncertainty in estimation of k^+ and k^- additional experimental observations on a longer time interval must be taken as well as the experiments with a different set of initial conditions need to be performed.

The fitted values of rate constants are small: $k^+ \ll 1$, $k^- \ll 1$; the distribution of k^- includes moderate values as well (we recall that we compare numerical values of the non-dimensionalized parameters with the characteristic length of the non-dimensionalized time interval of interest that is equal to 1). The reason behind the uncertainty in k^- estimation from the available data is easy to explain: if k^+ is small (i.e., the forward reaction is slow), then w is practically not produced during this reaction over characteristic time interval of interest, and thus, the actual rate of the reverse reaction $k^- \cdot w$ is small for k^- that are either small or moderate (with w being small). We may arrive at the following two reduced models that turn out to be equivalent with respect to the data under consideration, but that can be distinguished if additional experimental observations are collected.

Model case 1. First we consider the case where both $k^+ \ll 1$, $k^- \ll 1$. This means that the characteristic times associated with the forward and reverse binding reactions are long compared to the characteristic time interval: $1/k^+ \gg 1$, $1/k^- \gg 1$. Let us introduce a small parameter $0 < \varepsilon \ll 1$, so that the new re-scaled non-dimensional rate constants \hat{k}^+ , \hat{k}^- become moderate:

$$\hat{k}^+ = k^+/\varepsilon = O(1), \quad \hat{k}^- = k^-/\varepsilon = O(1). \quad (20)$$

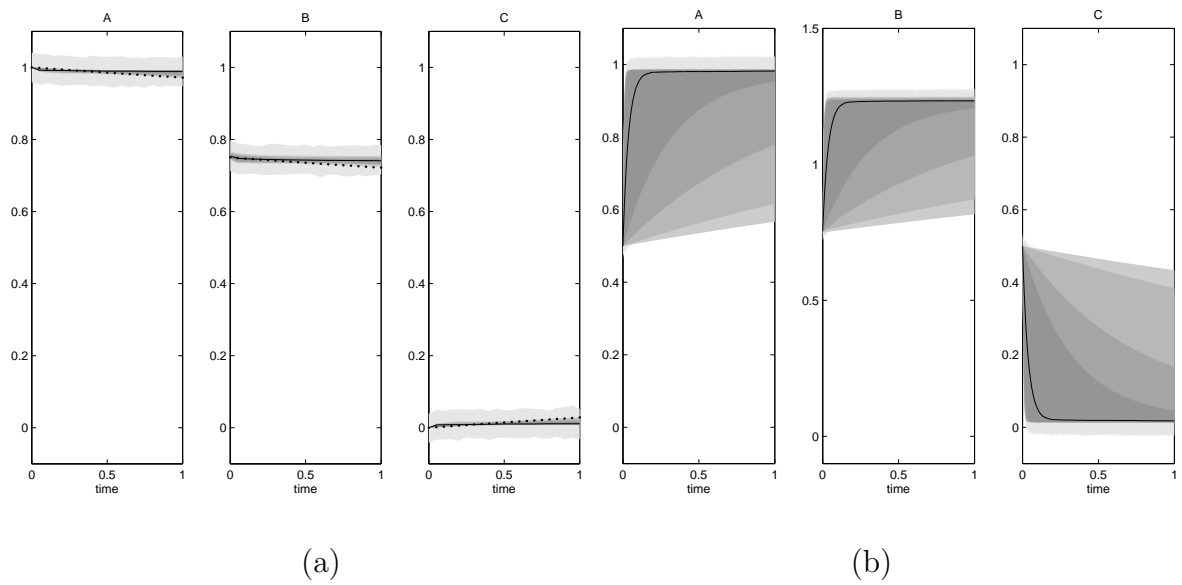


Figure 9: Case of slow reactions. Solid curves approximate the medians of model solutions obtained for 500 parameter value sets sampled from the chain. Model predictions: 95% confidence bands for means (dark grey) and 95% prediction bands for outcomes of single experiments (light grey). (a) Prediction for the original model with initial conditions: $u(0) = 1$, $v(0)$ is sampled from the chain, $w(0) = 0$. Dots correspond to data points. (b) Prediction for the model with initial conditions: $u(0) = 0.5$, $v(0)$ is sampled from the chain, $w(0) = 0.5 \neq 0$.

Once again, the fact that some parameters were small (or large) in the original non-dimensional model formulation will be shown explicitly using the small parameter factor. After substitution of (20) into the original model the new re-scaled model can be written as follows:

$$\frac{du}{dt} = -\varepsilon \cdot \hat{k}^+ \cdot u \cdot (u - 1 + v^*) + \varepsilon \cdot \hat{k}^- \cdot (1 - u), \quad 0 \leq t \leq 1; \quad (21)$$

$$u(0) = 1; \quad (22)$$

with two additional relations (11), (12).

The above model formulation represents a classical example of a *regularly perturbed problem*. The reduction of (11), (12), (21), and (22) on the closed interval $0 \leq t \leq 1$ may be performed by setting $\varepsilon = 0$ (see, e.g., Vasil'eva et al. [5] and Appendix 1). We arrive at a new model formulation:

$$\frac{d\bar{u}}{dt} = 0, \quad \bar{u}(0) = 1, \quad \bar{v} = \bar{u} - 1 + v^*, \quad \bar{w} = 1 - \bar{u}; \quad 0 \leq t \leq 1. \quad (23)$$

Here once again we use notation \bar{u} , \bar{v} , and \bar{w} to denote the approximations to u , v , and w now in the closed interval $0 \leq t \leq 1$.

The new *reduced* model (23) after integration of the differential relation with corresponding initial condition can be re-written as

$$\bar{u}(t) = 1, \quad \bar{v}(t) = v^*, \quad \bar{w}(t) = 0; \quad 0 \leq t \leq 1. \quad (24)$$

it contains not three but only one parameter: v^* . If we re-fit the reduced model containing this one parameter to the available data and use MCMC Metropolis algorithm for determining the distribution for this parameter, we will get: the estimated value obtained using MATLAB minimization software is $v^* = 0.7436$; the mean of the MCMC chain is $v^* = 0.74433$; estimated standard deviation is $\sigma_{v^*} = 0.0049842$, which is about 0.7% of the estimated value. Evidently, the value of v^* is estimated reliably.

Model case 2. Next we consider the case where $k^+ \ll 1$ is small, and $k^- = O(1)$ is moderate. This means that the characteristic time associated with the forward binding reaction is long *compared to the characteristic time interval*: $1/k^+ \gg 1$. Let us introduce a small parameter $0 < \varepsilon \ll 1$, so that the new re-scaled non-dimensional rate constant \hat{k}^+ becomes moderate:

$$\hat{k}^+ = k^+/\varepsilon = O(1). \quad (25)$$

After substitution of (25) in the original model the new re-scaled model can be written as follows:

$$\frac{du}{dt} = -\varepsilon \cdot \hat{k}^+ \cdot u \cdot (u - 1 + v^*) + k^- \cdot (1 - u), \quad 0 \leq t \leq 1; \quad (26)$$

$$u(0) = 1; \quad (27)$$

with two additional relations (11), (12).

Once again we arrive at a *regularly perturbed problem*. Same as earlier, the reduction of (11), (12), (26), and (27) on the closed interval $0 \leq t \leq 1$ may be performed by setting $\varepsilon = 0$ (see Appendix 1). We get a new model formulation:

$$\frac{d\bar{u}}{dt} = k^- \cdot (1 - u), \quad \bar{u}(0) = 1, \quad \bar{v} = \bar{u} - 1 + v^*, \quad \bar{w} = 1 - \bar{u}; \quad 0 \leq t \leq 1. \quad (28)$$

We use notation \bar{u} , \bar{v} , and \bar{w} to denote the approximations to u , v , and w in the closed interval $0 \leq t \leq 1$.

The new *reduced* model (28) that after integration of the differential equation with corresponding initial condition can be re-written as

$$\bar{u}(t) = 1, \quad \bar{v}(t) = v^*, \quad \bar{w}(t) = 0; \quad 0 \leq t \leq 1. \quad (29)$$

contains only one parameter: v^* (same as in the previous model reduction case).

To reliably identify moderate k^- one has to collect additional experimental data on the original time interval of interest for the situation where the initial condition for w is not zero.

Appendix 1: Asymptotic reduction of models formulated in terms of regularly/singularly perturbed ordinary differential equations

Here we briefly review some classical results related to asymptotic reduction of models formulated in terms of regularly and singularly perturbed differential equations; for a more detailed discussion see [5]. To simplify the explanation, we only discuss the regular perturbation case for a scalar differential equation, and we discuss the singular perturbation case for a system of two equations.

Regular perturbations. Consider the following problem:

$$\frac{du}{dt} = f(u, t, \varepsilon), \quad 0 \leq t \leq T; \quad (30)$$

$$u(0, \varepsilon) = u^*, \quad (31)$$

where $u(t, \varepsilon)$ is the dependent variable (unknown solution of a differential equation); $0 < \varepsilon \ll 1$ is a small parameter; $T = \text{const}$; $u^* = \text{const}$ is the initial condition for u ; $f(u, t, \varepsilon)$ is a known *once continuously differentiable function of its arguments*.

The uniform in $0 \leq t \leq T$ asymptotic approximation of the solution of (30), (31) may be constructed as a series expansion in the powers of ε . To obtain the reduced model for (30), (31), we may only consider the leading order approximation. We write:

$$u(t, \varepsilon) = \bar{u}(t) + O(\varepsilon), \quad (32)$$

where $\bar{u}(t)$ is the leading order term of asymptotic approximation for which the reduced model will be formulated. Notation $O(\varepsilon)$ has the following meaning: we say that a scalar

α satisfies $\alpha = O(\varepsilon)$ if there exists a constant $C > 0$ such that $|\alpha| \leq C \cdot \varepsilon$ for sufficiently small ε .

Substituting (32) into (30), (31), and retaining only the terms that do not contain ε , we obtain the problem for $\bar{u}(t)$,

$$\frac{d\bar{u}}{dt} = f(\bar{u}, t, 0), \quad 0 \leq t \leq T; \quad (33)$$

$$\bar{u}(0) = u^*, \quad (34)$$

that has a unique solution on $[0, T]$.

The solution \bar{u} of the *reduced model* (33), (34) approximates the solution of the original *regularly perturbed model* *uniformly* on the interval $0 \leq t \leq T$, i.e.,

$$\lim_{\varepsilon \rightarrow 0} u(t, \varepsilon) = \bar{u}(t), \quad 0 \leq t \leq T. \quad (35)$$

Singular perturbations. Now consider the following problem:

$$\varepsilon \frac{du}{dt} = f(u, z, t, \varepsilon), \quad (36)$$

$$\frac{dz}{dt} = g(u, z, t, \varepsilon),$$

$$0 \leq t \leq T;$$

$$u(0, \varepsilon) = u^*, \quad z(0, \varepsilon) = z^*. \quad (37)$$

Here $u(t, \varepsilon)$ and $z(t, \varepsilon)$ are the dependent variables (unknown solutions of differential equation system); $0 < \varepsilon \ll 1$ is a small parameter; $T = \text{const}$; $u^* = \text{const}$ and $z^* = \text{const}$ are the initial condition for u and z ; $f(u, z, t, \varepsilon)$ and $g(u, z, t, \varepsilon)$ are known *once continuously differentiable functions of their arguments*.

Consider a degenerate system obtained from (36) by setting $\varepsilon = 0$:

$$0 = f(\bar{u}, \bar{z}, t, 0), \quad (38)$$

$$\frac{d\bar{z}}{dt} = g(\bar{u}, \bar{z}, t, 0).$$

Condition 1. Let the first equation in (38) have an isolated root with respect to \bar{u} on the interval $0 \leq t \leq T$:

$$\bar{u}(\bar{z}, t) = f^{-1}(\bar{z}, t). \quad (39)$$

We note that such root may be non-unique; the final choice of the correct root will be specified below.

Let us substitute (39) into the second equation of (38),

$$\frac{d\bar{z}}{dt} = g(\bar{u}(\bar{z}, t), \bar{z}, t, 0), \quad (40)$$

and assume that a solution of (40) with initial condition

$$z(0, \varepsilon) = z^* \quad (41)$$

exists for all $0 \leq t \leq T$.

Next, consider the associated equation for $U(\tau)$ (compare with the first equation of (36)):

$$\frac{dU}{d\tau} = f(U, \bar{z}, t, 0), \quad (42)$$

where \bar{z} and t are considered to be parameters.

Condition 2. Let

$$\frac{\partial f}{\partial U}(\bar{u}(\bar{z}, t), \bar{z}, t, 0) < 0 \quad (43)$$

uniformly in (\bar{z}, t) , i.e., the steady state $U = \bar{u}(\bar{z}, t)$ of (42) is asymptotically stable in the sense of Lyapunov.

In the case of multiple roots (39) we only consider those satisfying (43).

Finally, consider (42) for $t = 0$ and $\bar{z} = z^*$:

$$\frac{dU}{d\tau} = f(U, z^*, 0, 0), \quad (44)$$

with initial condition

$$U(0) = u^*. \quad (45)$$

Condition 3. Let solution $U(\tau)$ of (44), (45) exists for $\tau \geq 0$ and approaches stationary point $U = \bar{u}(z^*, 0)$ as $\tau \rightarrow \infty$.

We say that the initial condition (44) belongs to the *domain of attraction* of the stationary point $U = \bar{u}(z^*, 0)$. Conditions 2 and 3 allow us to choose the unique roots among those given by (39).

Tikhonov's theorem on passage to the limit (see, e.g., [5]) states that under Conditions 1 – 3 the *singularly perturbed problem* (36), (37) has a unique solution and

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} u(t, \varepsilon) &= \bar{u}(t), & 0 < t \leq T; \\ \lim_{\varepsilon \rightarrow 0} z(t, \varepsilon) &= \bar{z}(t), & 0 \leq t \leq T. \end{aligned} \quad (46)$$

We note that the first relation in (46) holds on the open interval that does not include $t = 0$, and thus the original model (36), (37) is singularly perturbed (since \bar{u} does not satisfy the original initial condition: $\bar{u}(0) \neq u^*$). Corresponding *reduced model* is given by the problem (40), (40) for \bar{z} , with expression (39) for \bar{u} .

Appendix 2: MCMC basics

In the Bayesian approach the unknown parameter vector is interpreted as a random variable. The aim of the analysis is to find its distribution. Before experimental data is

available the parameter θ has a *prior distribution* $p(\theta)$. The observations y update the distribution $p(\theta)$ to the *posterior distribution* by the *Bayes formula*

$$\pi(\theta) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta) d\theta}, \quad (47)$$

Here $p(y|\theta)$ is the *likelihood function* that gives the likelihood of data y for given parameter value θ . The *posterior distribution* $\pi(\theta) = p(\theta|y)$ gives the probability distribution of parameter values, given the measured data y . The integral $\int p(y|\theta)p(\theta) d\theta$ is needed as the normalizing constant, to ensure that π indeed is a probability measure, with total measure equal to one, $\int_{\theta} \pi(\theta) d\theta = 1$.

In the usual settings the parameter vector θ and data y are connected by a model $y = f(x; \theta) + \epsilon$, where the experimental error $\epsilon \sim N(0, \sigma^2 I)$, i.e., in all experiments the measurement noise is Normally distributed, independent and with standard deviation of size σ . It is not difficult to see that then

$$p(y|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^n (y_i - f(x_i; \theta))^2 / 2\sigma^2}.$$

So, we arrive at the familiar least squares function: maximizing the likelihood function turns out to be equivalent to minimizing the residual sum of squares.

In principle, the Bayes formula solves the estimation problem in a fully probabilistic sense: we find the peak, the maximum a posteriori (MAP) point, of the parameter distribution. Then we determine a required portion of the probability mass (typically some 95% or 99% of the mass) around it. However, we face the problem of how to define the a priori distribution, and how to calculate the integral of the normalizing constant. Often, only a 'flat' prior is used, that is, a uniform distribution that only defines physically possible lower and upper bounds for each parameter. However, the integration of the normalizing constant often is a formidable task, even for only moderately high number of parameters in a nonlinear model. So, a direct application of the Bayes formula is intractable for all but trivial nonlinear cases. The MCMC methods provide a tool to handle this problem. They generate a sequence of parameter values $\theta_1, \theta_2, \dots, \theta_N$, whose empirical distribution approximates the true posterior distribution for large enough sample size N .

The trick here is that we do not know the distribution from which to sample, but we still can generate samples from it. Instead of sampling from the true distribution, we only may sample from an artificial *proposal* distribution. Combining the sampling with a simple accept/reject procedure, the posterior can be correctly approximated. The simplest MCMC method is the *Metropolis algorithm*:

- Initialize by choosing a starting point θ_1 .
- Choose a new candidate $\hat{\theta}$ from a suitable **proposal** distribution $q(\cdot|\theta_n)$ that may depend on the previous point of the chain.
- **Accept** the candidate with probability

$$\alpha(\theta_n, \hat{\theta}) = \min \left(1, \frac{\pi(\hat{\theta})}{\pi(\theta_n)} \right).$$

If rejected, repeat the previous point in the chain. Go back to item 2.

So, points with $\pi(\hat{\theta}) > \pi(\theta_n)$, i.e., steps 'uphill', are always accepted. But also points with $\pi(\hat{\theta}) < \pi(\theta_n)$, i.e., steps 'downhill', may be accepted, with probability that is given by the *ratio* of the π values. In practice, this is done by generating a uniformly distributed random number $u \in [0, 1]$ and accepting $\hat{\theta}$ if $u \leq \pi(\hat{\theta})/\pi(\theta_i)$. Note that only the ratios of π at consecutive points are needed, so the main difficulty is omitted: the calculation of the normalizing constant is not needed since the constant cancels out!

However, the choice of the proposal distribution may still pose a problem. It should be chosen so that the 'sizes' of the proposal q and target distributions suitably match. This often may be difficult to achieve. An unsuitable proposal leads to inefficient sampling, typically due to

- the proposal being too large. Then the new candidates mostly miss the essential region of π ; they are chosen at points where $\pi \simeq 0$ and only rarely accepted.
- the proposal being too small. The new candidates mostly are accepted, but from a small neighborhood of the previous point. So, the chain moves only slowly, and may not cover the target π in finite number of steps.

For simple cases, the proposal might be relatively easy to find by some hand-tuning. However, the 'size' of the proposal distribution is not a sufficient specification. In higher dimensions, especially, the shape and orientation of the proposal are crucial. The most typical proposal is a multi-dimensional Gaussian (Normal) distribution. In the *random walk* version, the center point of the Gaussian proposal is chosen to be the current point of the chain. The task then is to find a covariance matrix that produces efficient sampling. Several efficient adaptive methods have been recently proposed (see, e.g., ([3]),([4])); this topic remains an active research field.

Acknowledgement. The work of Leonid Kalachev in this project was supported by NIH Grant P20RR015583 from the COBRE Program of the National Center for Research Resources.

References

- [1] Yonathan Bard. Nonlinear Parameter Estimation. Academic Press (1974), ISBN: 978-0120782505.
- [2] A. Gelman, J.B. Carlin. Bayesian Data Analysis, Second Edition. Chapman & Hall (2004). ISBN: 978-1584883883.
- [3] Heikki Haario, Eero Saksman, Johanna Tamminen. An Adaptive Metropolis Algorithm, Bernoulli, Vol. 7, Number 2 (2001), 223-242.
- [4] Haario H., M. Laine, A. Mira, E. Saksman: *DRAM: Efficient adaptive MCMC* . Stat. Comput. 16, pp. 339–354, ISSN 0960-3174, 2006.

- [5] A.B. Vasil'eva, A.B. Butuzov and L.V. Kalachev: *The Boundary Function Method for Singular Perturbation Problems*, Philadelphia: SIAM, 1995.
- [6] J. Kevorkian and J.D.Cole, *Singular Perturbation Methods in Applied Mathematics*, Berlin-Heidelberg-New York, Springer-Verlag, 1981.
- [7] R.E. O'Malley, *Singular Perturbations Methods for Ordinary Differential Equations*, Berlin-Heidelberg-New York, Springer-Verlag, 1991.
- [8] Haario H., Kalachev L. and M. Laine: *Reduced Models for Algae Growth*. Bulletin of Math. Biology, Vol. 71, 7, 1626–1648, 2009.