Inverse problems and computational cell metabolic models: a statistical approach

To cite this article: D Calvetti and E Somersalo 2008 J. Phys.: Conf. Ser. 124 012003

View the article online for updates and enhancements.
Inverse problems and computational cell metabolic models: a statistical approach

Daniela Calvetti¹, Erkki Somersalo²

¹Department of Mathematics, Case Western Reserve University, 10900 Euclid Avenue, OH 44106, USA, ²Department of Mathematics, Helsinki University of Technology, PO Box 1100, FIN-02015 TKK, Finland

E-mail: daniela.calvetti@case.edu, erkki.somersalo@hut.fi

Abstract. In this article, we give an overview of the Bayesian modelling of metabolic systems at the cellular and subcellular level. The models are based on detailed description of key biochemical reactions occurring in tissue, which may in turn be compartmentalized into cytosol and mitochondria, and of transports between the compartments. The classical deterministic approach which models metabolic systems as dynamical systems with Michaelis-Menten kinetics, is replaced by a stochastic extension where the model parameters are interpreted as random variables with an appropriate probability density. The inverse problem of cell metabolism in this setting consists of estimating the density of the model parameters. After discussing some possible approaches to solving the problem, we address the issue of how to assess the reliability of the predictions of a stochastic model by proposing an output analysis in terms of model uncertainties. Visualization modalities for organizing the large amount of information provided by the Bayesian dynamic sensitivity analysis are also illustrated.

1. Introduction

The word metabolism seems to be ubiquitous whenever the important issues of human health are discussed. The interest in understanding human cellular metabolism dates back a few thousand years, when already the Greek philosophers linked the state of health of the human body, and of the human mind, to the foods which were being consumed and to their internal processing by the body. Today, with the epidemic proportions that a metabolism-related disease like type II diabetes is taking, the interest in understanding human cell metabolism is stronger than ever. In fact, while it is known and accepted that several diseases can be blamed on a faulty component of the metabolic system [12], for a large class of pathologies there is still not a definite answer about the mechanics at the cellular and subcellular level [22, 23]. In addition to type II diabetes, which is the new plague of the developed world and increasingly also of the emerging societies, it has been suggested that the Alzheimer’s and the Parkinson’s diseases, whose incidence is on the rise, may, too, be explained in terms of perturbed metabolic functioning [18, 21]. A good knowledge of how the metabolic processes adapt to an emergency or to a crisis can be useful in designing clinical procedures. For example, the critical time after which the damage caused by a heart attack becomes irreversible can be estimated in terms of the metabolic changes occurring in the myocardial tissue when the supply of metabolic substrates is severely limited by a blockage in the blood flow. The question of what should be administered to patients after cerebral ischemia, to minimize, or at the very least contain, the re-perfusion injury, requires a deep understanding...
of the metabolic processes in the brain. Since the ramifications of metabolic disfunctions reach
very far, from genetics [24] to cancer management [11], the study of cellular metabolic processes
is a core area in life science.

Most of the literature about human cellular metabolism has been compiled by researchers
with a solid biochemical, biophysical and physiological background. While the importance of
experimental results cannot be overemphasized, the need for a key to interpret the results has
come become clear. Moreover, as the instrumentation capabilities increase, the collection of data
needs to be targeted. Computational models can be very helpful in this respect, because they
can provide in silico results which are impossible, or unethical, to gather in vivo (in situ), that
can be used to support or contrast various hypotheses about metabolic systems. Mathematical
models have been utilized in several different applications, and in recent years the need for
computational mathematical models has been advocated by computational systems biology as
a necessary step to understand some of the complex metabolic processes suspected of being
responsible for key human functions.

When designing a mathematical model it is important to have a good understanding of
the underlying biological application. The transformation of a mathematical model into
a computational one however cannot happen without a robust and efficient computational
methodology which allows in silico simulation of realistic experiments and a maximally unbiased
interpretation of the results.

A crucial input in the design of a computational model for cellular metabolism is the
type investigations targeted, that is whether it will applied to validate hypotheses about the
functioning of a metabolic system under a given set of conditions, or to predict the behavior
of biochemical pathways impossible to measure experimentally. The two functions are different
but not completely separate: in fact, the validation procedure is a necessary step to ensure that
the computational model is able to simulate adequately the actual system producing output in
agreement with measured data, prior to using the model for predictive purposes.

The variability of measured biological data is a generally well accepted fact, re-emphasized
by the routine repetition of laboratory tests prescribed by physicians when the reported values
are outside the “normal” range. Often it is by collecting a larger sample that the results which
might be of concern can be treated as outliers. The intrinsic stochastic nature of biological
systems makes it very natural to embed the design and analysis of cellular metabolic systems
into a stochastic framework. The success of the methodology that we have developed over the
past few years [8, 1, 2, 5, 13, 19] at advancing the understanding of biological systems has been
the main motivation to make it as much as possible accessible to researchers in the biological and
medical field, and to continue working on improving its robustness and computational efficiency.

In this paper we will give an overview of how to set up a statistical framework for combining
the biochemistry behind cellular metabolic processes, mathematical formulations and robust
and efficient numerical methods into a powerful methodology for solving some of the important
inverse problems of cellular metabolism. In addition to the previously published applications
to skeletal muscle metabolism during ischemia [8, 13] and exercise [3], myocardial metabolism
under ischemia [1, 2] metabolic zonation of liver gluconeogenesis [5], and brain energy metabolism
[19], details about applications of this methodology to the investigation of the role of ketone
bodies in the energetics of brain metabolism and to the hepatic gluconeogenesis can be found
in [7, 6] in this issue. A more detailed account of the computational challenges that had to
be overcome, and of those which still lie ahead can be found in [?], also in this issue. The
paper is organized as follows. In Section 2 we briefly outline multi-compartment models for
cellular metabolism, pointing out the reasons for compartmentalization and the increase in
computational complexity coming with each additional compartmentalization. Section 3 is
concerned with the dynamic mass balance equations which govern our models, introduces the
Michaelis–Menten kinetic expressions for the biochemical reaction fluxes and transport rates and
the convective term associated with the blood flow. The transition of our modelling methodology from deterministic to stochastic is explained in Section 4. Our statistical model comprises elements from both frequentists’ population approach and Bayesian subjective probability. The genesis of two inverse problems of cellular metabolism, and their importance in the biological and medical field are the topics of Section 5. In Section 6 we explain how to analyze the output of our computational experiments within the statistical framework, and we introduce credibility envelopes as a tool for a preliminary reliability assessment for the results of in silico experiments. The dynamic nature of sensitivity analysis for kinetic models and its probabilistic interpretation are discussed in Section 7. A discussion of the overall approach and a few concluding remarks can be found in Section 8.

2. Multi-compartmentalization

The human body consists of many different organs with different needs and capabilities. The particular tasks the organs perform are reflected in their specific biochemical functioning. To understand the energy metabolism of a given organ, one needs to have a specific cellular model for the key metabolic processes, reactions and transports that characterize it.

The spatially lumped cellular metabolic model for an organ, or part thereof, is based on a compartment model, where each compartment is justified by a physiological counterpart [8]. The crudest compartmentalization is to subdivide the organ into two interacting compartments, the blood and the tissue. Often, such a rough division does not have enough resolution of the important processes. The heart, for example, which is responsible for keeping the blood flowing through the vascular system to distribute nutrients to the various organs while collecting the waste products, needs to continuously produce the energy required to make its rhythmic contractions possible. In view of the high oxidative cardiac metabolism, it is quite natural to include in the model for myocardial metabolism details about the organelles which are most responsible for it, namely the mitochondria [1]. More complex organs, like the brain, include different cell types whose interactions are quite interesting and intriguing. Therefore if the model is used to investigate functions which are regulated by the concerted action of different cell types, they each need to be modelled separately, thus adding a new layer of compartmentalization, and the important details in each cell type dictate how many subcompartments each layer contains. Our most recent model of brain energetics [19] accounts for astrocytes and neurons separately. Since for each cell type it is important to distinguish reactions which occur in cytosol from those taking place in mitochondria, the minimum number of compartments for a model of this type is four. Neurons are subdivided into glutamatergic and GABAergic, depending on which neurotransmitter they release, and since their energetic demands and biochemical processes are significantly different, a realistic model of brain energetics requires an additional layer of compartmentalization to distinguish between the two neuron types, raising the number of compartments to six. The blood, which is responsible for delivering substrates to the cells and for removing the end products of metabolic processing, is accounted for in a separate compartment. The exchange of substrates is sometimes more complex; for example, since the brain is protected by the blood brain barrier, it would be reasonable to model the intercellular space as a separate compartment. Thus, more detailed models of cerebral brain energetics should include at least eight separate compartments. The addition of each new compartment increases substantially the overall complexity of the model, and raises considerably the computational challenges, explaining why the metabolic models in the literature are rather coarsely compartmentalized.

Many metabolic problems can be studied using a spatially homogeneous model, which implicitly assumes that the spatial location does not affect the metabolic processes. In some organs, for example the liver, it is known that different gene expressions of the enzymes along the sinusoids of the hepatic lobules occur at different spatial locations, and one of the main questions is how this zonation of the enzymes affects liver functions. Thus, in order to perform
computational experiments aimed at studying the effect of zonation, it is necessary to design a spatially distributed model, able to account for differences in the metabolic functions at different spatial locations. The spatially distributed model for liver metabolism proposed in [10, 5], which discretizes a liver sinusoid into several separate tissue compartments, accounts for each biochemical species present in each tissue compartment separately, thus quickly leading to a system of very large dimensions.

In summary, we can view a metabolic system as a collection of different compartments, $C_1, C_2, \ldots, C_n$, a corresponding collection of state vectors $C^1, C^2, \ldots, C^n$ containing the concentrations of the biochemical species tracked in the respective compartment and a collection of neighborhood lists $N^1, N^2, \ldots, N^n$, tracking the trafficking of biochemical species among the various compartments.

In the most general setting, we expect the concentrations of the different species to change in time because of biochemical reactions and transmembrane transports, and therefore the reactions fluxes and transport rates to change accordingly. It is the goal of a dynamic model of cellular metabolism to be able to reliably track these changes, providing the necessary information to infer about the metabolic processes.

We conclude this section illustrating the biochemical pathways associated with a few different types of multi-compartment models.

**Example 1**: The first example is a two-compartment model of the skeletal muscle metabolism, which distinguishes only blood and tissue compartments, thus keeping the number of transports and biochemical species to a minimum. The number of biochemical species tracked in the two compartments is 39, while a total of 26 reaction fluxes and 11 transport rates are included in the model [8, 13]. The biochemical pathways summarizing the metabolic system is shown in Figure 1.

![Figure 1. Biochemical pathway of a two compartment model of skeletal muscle metabolism. The nodes in the diagram indicate a biochemical species, and the arrows indicate either biochemical reactions or transports between the compartments.](image)

**Example 2**: In the three-compartment model of myocardial metabolism considered in this example the tissue is separated into cytosol and mitochondria, to follow more carefully the redox states in the two cellular compartments during an ischemic episode. An additional sub-compartmentalization of the glycolytic pathways is introduced, to include the information that most of the biochemical reactions participating in glycolysis occur very close to the membrane.
separating cytosol from the blood compartment. The biochemical pathway corresponding to this model is displayed in Figure 2. The model was used to study the dynamics of the myocardial metabolism during an ischemic episode [1, 2].

Figure 2. Biochemical pathway of a myocardial cellular metabolic model used to simulate an ischemic event.

Example 3: This example considers a five-compartment cellular brain metabolism model. The biochemical pathway, shown in Figure 3, illustrates how, even lumping together the blood and intercellular space into a well-mixed compartment, the number of biochemical reactions and transport rates increases tremendously. An investigation of brain energy metabolism at high and low neuronal activity performed on a variant of this model has been presented in [19]. An extension which includes the metabolic pathways of the ketone bodies was used for the analysis of brain energy metabolism under ketosis and is presented in this same issue [7].

Example 4: The spatially distributed model of the hepatic metabolism proposed in [5, 10] subdivides the tissue along an hepatic sinusoid within a hepatic globule into several independent subcompartments, each one equipped with its own metabolic pathway. One such pathways is displayed in Figure 4. In view of the high complexity of this model, coming from the spatial distribution, the single metabolic units are kept fairly simple.

3. Convection, reactions and transports

Once the metabolic network is set up, we first write the dynamic system for substrate concentrations following a classic paradigm of mass balance. Our prototype equation of spatially lumped cellular metabolism is of the form:

$$V^j \frac{dC^j}{dt} = K^j + \sum_{k \in N_j} T^{jk} + R^j, \quad 1 \leq j \leq n.$$  

Here, $V^j$ is the volume of the compartment and $C^j$ is a vector containing the metabolite concentrations in the compartment $C_j$. Since biochemical reactions only occur in tissue, the reaction term is typically absent from the blood equations, except for some tissues, for example, adipose tissue, where it may be convenient to model the esterification of glycerol into triglycerides.
Figure 3. Biochemical pathways of a five compartment model of cellular brain metabolism, where the neurons and the glial cells, astrocytes, are modelled separately.

Figure 4. Biochemical pathways of a prototypical hepatocyte for a model used to study liver gluconeogenesis.

in blood as a reaction. Conversely, the convection term is related directly to the blood flow and is therefore not present in the equations for the cellular compartments.

To specify the various terms in the model equation above, consider the prototype equation in the simplest case of two compartments (Example 1 above). In the blood compartment, the
mass balance equation is written as

\[ V_b \frac{dC_b}{dt} = \frac{Q(t)}{F} (C_a(t) - C_b(t)) + J_{c\rightarrow b} - J_{b\rightarrow c}. \]  

(1)

\( C_b \) is the vector of the blood concentrations of the biochemical species tracked in blood and \( C_a \) is the arterial concentration vector. The blood flow coefficient \( Q(t) \) is time dependent because the blood flow may change during the course of an experiment, and the coefficient \( F \) is the mixing ratio of the arterial and venous concentrations from which we obtain \( C_b \). The transport term is defined as the difference between the transport rates of the species from cell to blood (\( J_{c\rightarrow b} \)) and from blood to cell (\( J_{b\rightarrow c} \)). Finally, \( V_b \) is the volume of the blood compartment.

In the same way, the prototype equation for a generic tissue compartment is

\[ V_c \frac{dC_c}{dt} = S\Phi + M \left( J_{b\rightarrow c} - J_{c\rightarrow b} \right), \]

(2)

where the matrix \( M \) identifies the species exchanged between compartments and matches the corresponding transport rates and concentration equations, \( S \) is the stoichiometric matrix, whose entry \( s_{i,j} \) indicates the number of moles of species \( i \) produced (\( s_{i,j} > 0 \)) or consumed (\( s_{i,j} < 0 \)) in the reaction \( j \), \( \Phi \) is the vector of all reaction fluxes occurring in the compartment and \( V_c \) is the tissue compartment volume. We remark that the two matrices \( M \) and \( S \) are the interface between the mathematical formulation of the model and the biochemical properties of the system. A correct stoichiometry is of utmost importance when setting up the model, and while the stoichiometric relations for many reactions are well known, in some case the information may be difficult to get, in particular when several reactions are lumped into a single one. Furthermore, in some multicompartment models it is not always clear where some of the biochemical reactions take place, and this can be in itself a topic of investigation.

To emphasize the difference that a compartmentalization makes in the problem setup and the different flavors that a study of metabolic processes can have, it suffices to think that, while in the myocardial metabolic model of Example 2 the Citric Acid Cycle was quite detailed and located in the mitochondria compartment, in the hepatic gluconeogenesis model of Example 4, the entire Citric Acid Cycle was represented by a single biochemical reaction, and no separation was made in the cellular tissue between cytosol and mitochondria. The lumping process reduces the complexity of the model, but also makes impossible to track any of the changes in the Citric Acid Cycle intermediates concentrations.

3.1. Parametric model for reaction fluxes

At steady state the unknowns of primary interest, reaction fluxes and transport rates, are assumed to remain constant. In a fully dynamic simulation, on the other hand, where the changes in the values of the reaction fluxes reflect the changes of the concentrations of the biochemical species, it is necessary to choose suitable kinetic expressions. Since all biochemical reactions occurring in cells are enzymatic, unless there are compelling reasons to do otherwise, we will model the corresponding biochemical reaction fluxes in Michaelis-Menten form \([14]\). Let us consider a single substrate facilitated reaction of the form

\[ A + E \rightarrow B + \tilde{E} \]

where \( A, B \) are metabolites and \( E, \tilde{E} \) is a facilitator pair. The Michaelis-Menten form that model the corresponding reaction flux \( \Phi_{A\rightarrow B} \) is

\[ \Phi_{A\rightarrow B} = V_{\text{max}} \frac{P}{\mu + P} \frac{C_A}{K + C_A}, \quad P = \frac{C_E}{C_{\tilde{E}}}. \]
In the case of a bi-substrate facilitated reaction

\[ A_1 + A_2 + E \rightarrow B_1 + B_2 + \tilde{E} \]

we modify the previous Michaelis-Menten form to

\[ \Phi_{A_1+A_2 \rightarrow B_1+B_2} = V_{\text{max}} \frac{P}{\mu + P} \frac{C_{A_1}C_{A_2}}{K + C_{A_1}C_{A_2}}. \]

3.2. Parametric model of transport rates

Cell membranes consist of a double layer of phospholipid molecules, interspersed with channels, which allow the free passage of some substances, and restrict others. Molecules can cross a membrane via either passive or facilitated transport. In a simplified model of transmembrane transports, smaller molecules can be thought of as crossing the membrane through free diffusion, while larger ones typically require bonding to a carrier which facilitates their passage [17]. Therefore we model the free diffusive transport of the species \( A \) from compartment \( x \) to compartment \( y \) in the form

\[ J_{A,x \rightarrow y} = \lambda (C_{A,x} - \sigma C_{A,y}), \]

where \( \lambda \) is the diffusion coefficient and \( \sigma \) is a correcting factor taking care of the uneven distribution of the substrate within the cell or organelle, while the carrier facilitated transport are expressed in a Michaelis-Menten form analogous to that used for biochemical reaction fluxes,

\[ J_{A,x \rightarrow y} = T \frac{C_{A,x}}{M + C_{A,x}}. \]

We remark that the system parameters \( V_{\text{max}}, K, \mu, T, M, \lambda \) and \( \sigma \) are reaction or transport specific, hence each additional reaction or transport increases by two or three the number of parameters whose values need to be assigned in order for the system to be specified. Furthermore, since the dynamic mass balance equations are in the form of an initial value problem, the initial values of the tracked concentrations also need to be assigned.

Collecting all system parameters into the vector \( \Theta \),

\[ \Theta = [V_{\text{max}}, K, \mu, T, M, \lambda, \sigma], \]

and all initial concentrations into the vector \( C_0 \), we have that

\[ C(t) = C(\Theta, C_0, t). \]

This traditional deterministic model is what most of the modelling literature deals with; for details see [25, 10].

4. From deterministic to stochastic models

So far, the description of the metabolic system has followed the classical deterministic paradigm, connecting cells, blood and interstitial fluid via a deterministic dynamical system. The parameters which specify the system, e.g. the Michaelis-Menten constants, summarize the influence of various factors, for example enzymatic activities which, in turn, are gene expressions, and hence depend on regulatory hormones, with insulin the most prominent examples. This observation should make it clear that no deterministic model can explain extensively the metabolism over a population, or even for a single individual because of the intrinsically variable nature of the process. Rather, a more natural description is to specify a distribution of models.
Consider the deterministic dynamic model (3). Assuming that the model parameter vector \( \Theta \) and the initial value \( C_0 \) are random variables with the probability distribution \( \pi(\Theta, C_0) \), the model naturally induces a probability density also for \( C \). While deterministic predictive models produce only one single predicted output, **predictive statistical models** give the probability distribution of the concentrations. The difference is fundamental: the statistical output prediction seeks to explain the behavior of the system over the entire population, and it also accounts for the credibility of the output.

5. **Inverse problems of cellular metabolism**

Inverse problems are ubiquitous when studying cellular metabolism. The mystery of what happens to the food that we eat has fascinated scientist and philosophers alike for many centuries, and the desire to understand the connections between what we eat and what we are is a recurrent theme of research. While the big inverse problem of cellular metabolism may well be to understand what is affecting who we are, or differently put, which foods are responsible for certain illnesses and states of mind, at the biological and medical end the corresponding inverse problem can be stated as finding out how certain substrates configurations affect the internal functioning of our organs.

In our design of a methodology for modelling and simulating cellular metabolism, we have addressed two broad classes of inverse problems: finding the values of those reaction fluxes and transport rates which are the key players in the metabolic function at the cellular and subcellular level of a particular organ at steady state, and identifying a dynamic model capable of describing the changes in metabolites and intermediates over a time interval in which the concentrations of the biochemical species undergo some changes.

The two inverse problems pertain two different types of situations, and their complexity is quite different. At steady state case it is assumed that the unknowns of primary interest, reaction fluxes and transport rates, remain constant, hence they can be estimated once and for all. Thus the dimensions of this inverse problem are equal to the number of biochemical reactions and transmembrane transport rates included in the model.

In the fully dynamic setup, on the other hand, the concentrations of metabolites and intermediates can change in time and the inverse problem amounts to identifying a dynamic system where the number of parameters is proportional to the number of reaction fluxes and transmembrane transports considered, with a proportionality constant no smaller than 3 or 4.

The dimensionality of our cellular metabolism inverse problems is rather high because our models are fairly detailed, ranging from little less than one hundred for the steady state case, to several hundreds in the fully kinetic case. The scarcity of data, the high variability of the available measurements and the complexity of the underlying biochemistry and physiology makes the solution of these inverse problems rather challenging.

5.1. **Bayesian flux balance analysis of steady state models**

As indicated in Section 4, rather than seeking one solution as in deterministic inverse problem, the more natural approach is to seek a distribution, which quite naturally brings us to the statistical framework of inverse problems, and to the Bayesian one in particular.

The mass balance equations (1–2) can be compactified into a system of differential equations,

\[
\mathcal{V} \frac{d\mathbf{C}}{dt} = \mathbf{A} \mathbf{u},
\]

where \( \mathcal{V} \) is a diagonal matrix containing the volumes of the compartments, \( \mathbf{u} \) is a vector whose components are the reaction fluxes, transport rates and arterial and venous concentrations. Under the steady state assumption, the left side of (4) vanishes, and some of the components of \( \mathbf{u} \), for example the arterial concentrations, could be approximately known. We may then
decompose $\mathbf{u}$ in two parts, $\mathbf{u} = \mathbf{v} + \mathbf{w}$, where $\mathbf{w}$ contains the observed quantities in its non-vanishing components, while the non-observable ones are collected in $\mathbf{v}$. If we assume that the system is approximately at a steady state, we obtain the model

$$\mathbf{b} = A\mathbf{v} + \mathbf{e},$$

where $\mathbf{b} = -A\mathbf{w}$ contains the data, while $\mathbf{e}$ is a noise vector modelling the observation errors as well as the uncertainty about the system being indeed in the steady state. Following the Bayesian paradigm \[9, 16\], all variables are modelled as random variables. Assuming a probability density $\pi_{err}$ for the modelling error and measurement noise $\mathbf{e}$, we can express the likelihood in the form

$$\pi(\mathbf{b} \mid \mathbf{v}) = \pi_{err}(\mathbf{b} - A\mathbf{v}), \quad \mathbf{b} = \mathbf{b}_{\text{observed}},$$

where we have tacitly assumed statistical independence of $\mathbf{v}$ and $\mathbf{e}$. Generally, the number of observations is much smaller than the number of unknowns, which means that the likelihood $\mathbf{v} \mapsto \pi(\mathbf{b} \mid \mathbf{v})$ is an improper density. To have a useful model, we need to equip it with complementary information coming from the biological, biochemical and biomedical properties. In vitro measurements of enzymatic reactions provide estimates for upper bounds for some reactions and transports; thermodynamics dictates the prevailing direction of some reversible reactions under normal conditions; positivity constraints can be strengthened with stricter bound, assuring that some key reactions must occur, for example guaranteeing a certain level of neuronal activity; non-linear bound constraints can model allosteric regulation of certain enzymes and, finally, direct measurements or previously obtained estimates for some of the reactions may be available. All these information is encoded in the prior density, $\pi_{\text{prior}}(\mathbf{v})$, and Bayes’ formula leads to the posterior density

$$\pi_{\text{post}}(\mathbf{v}) \propto \pi_{\text{prior}}(\mathbf{v}) \pi_{err}(\mathbf{b} - A\mathbf{v}).$$

The classical flux balance analysis (FBA) seeks to find a single solution to the steady state equation $\mathbf{b} = A\mathbf{v}$, which is assumed to hold strictly, with the additional requirement that all the a priori constraints are respected \[15\]. This approach suffers from several shortcomings: first, it may be impossible to find a strict steady state, and to modify a complex metabolic network in a controlled way so as to allow a feasible solution may be a tremendous task. Second, the matrix $A$ may, and usually does, contain a significant null space, making any single solution rather unreliable. Optimization based selection criteria have been proposed in the literature to single out one solution among the many possible ones \[20\].

The Bayesian flux balance analysis (BFBA), on the other hand, simply explores the posterior density by using Markov Chain Monte Carlo (MCMC) techniques. This approach makes it possible to analyze the quasi-steady state of nearly any metabolic network. The inconsistency of the system, if present, usually shows up as slow convergence, while the non-uniqueness simply broadens the posterior densities, thus quantifying the degree of uncertainty in a palpable way \[13, 3, 4\].

5.2. Bayesian parameter identification of dynamic models

The steady state analysis does not yield probability distributions of the concentrations. To have a predictive model of concentrations, the distributions of dynamic parameters are needed. Evidently, the steady state data is not sufficient to identify the dynamic model parameters, and the addition of transient data information is required for this task. We proceed to estimate the distributions of the dynamic parameters and initial concentrations following the Bayesian paradigm. Assuming that a few concentrations, either in the blood or in the tissue, are measured at times $t_1 < t_2 < \ldots < t_k$, we may write an observation model

$$\mathbf{b}_j = P_j C_j(\Theta, C_0) + \mathbf{e}_j, \quad C_j(\Theta, C_0) = C(\Theta, C_0, t_j), \quad 1 \leq j \leq k,$$
where the matrix $P_j$ extracts from the model output the predicted values of the observed concentrations and $e_j$ is the measurement error at time $t_j$. More generally, the data may depend non-linearly on the concentrations. By stacking all the observations and the concentration vectors into a single vector, we get a model of the form

$$b = PC(\Theta, C_0) + e.$$  

As in the steady state case, we form the posterior density from the likelihood and prior,

$$\pi_{\text{post}}(\Theta, C_0) \propto \pi_{\text{prior}}(\Theta, C_0) \pi_{\text{err}}(b - PC(\Theta, C_0))$$

where the notation is consistent with the one introduced above.

Markov Chain Monte Carlo techniques are typically employed for the exploration of the posterior. Sampling techniques for this type of applications may be time consuming, since to test each proposal for acceptance it is necessary to solve systems of ordinary differential equations which, in view of the large range of time scales, tend to be very stiff. The long time required for the numerical solution of stiff systems is a great motivation for designing efficient sampling schemes. In [1, 8] we propose a sequential sampling scheme where an initial check for acceptance of the proposal is done using a steady state model, and only if this surrogate test is passed, the time propagation is performed. This scheme makes the rejection step quite efficient, speeding up the MCMC algorithm by orders of magnitude.

6. Bayesian output analysis and credibility envelopes

The solution of the dynamic Bayesian inverse problem is a probability distribution of the parameters identifying the model. Once this distribution is available, it is possible to perform posterior simulations to analyze the model predictions. Let $\pi_{\text{post}}(\Theta, C_0)$ denote the posterior distribution, and consider a prediction model for a time course of a quantity $\gamma$ that is a function of the concentration vector,

$$\gamma(t) = F(C(t)).$$

Typical quantities of interest include the redox state in the cytosol, measured by the concentration ratio $[\text{NAD}^+] / [\text{NADH}]$.

Estimating $\gamma$ with a deterministic predictive output system would amount to producing a single output curve from the computed estimate $(\Theta_{\text{est}}, C_{0,\text{est}})$ of the model parameters,

$$\gamma_{\text{est}}(t) = F(C(t, \Theta_{\text{est}}, C_{0,\text{est}})).$$

The natural question arising at this point is how reliable such a prediction is. The sensitivity of the output to perturbations of the parameters can be estimated by calculating the Jacobian of the right hand side. The Jacobian, however, might not take fully into account the information that the estimator is based on. In fact, it is possible that in the light of the measurements, perturbations in only a few directions in parameter space are possible, while others might be in conflict with the data and the prior information.

In the Bayesian interpretation, we may use the concept of conditioning, and define

$$\gamma(t \mid \Theta, C_0) = F(C(t, \Theta, C_0)), \quad (\Theta, C_0) \sim \pi_{\text{post}}(\Theta, C_0),$$

that is, each output curve is conditioned on the particular value of the parameter vector, and the parameter vector is distributed according to the posterior density. In our Bayesian posterior simulation, we generate a sample of output curves,

$$\{\gamma_1(t), \gamma_2(t), \ldots, \gamma_N(t)\}, \quad \gamma_j = \gamma(t \mid \Theta_j, C_{0,j}),$$
by first drawing a sample \( \{ (\Theta_j, C_{0,j}) \} \) from the posterior density. Notice that if MCMC methods have been used to explore the posterior, there is no need to produce this sample, since we already have it as part of the solution. The model uncertainty can then be expressed by plotting pointwise predictive output envelopes, that is, envelopes that contain a given percentage \( p\% \) of all the curves corresponding to all parameter vectors in the sample. The envelopes for the time courses of cytosol redox (left) and of the rate of oxidative phosphorylation in skeletal muscle (right) during a 20 minutes ischemic episode, are displayed in Figure 5.

![Figure 5](image-url)

**Figure 5.** Pointwise predictive output envelopes for the time courses of cytosol redox (left) and of the rate of oxidative phosphorylation in skeletal muscle (right) during a 20 minutes ischemic episode. The envelopes contain 50% (darker) and 90% (lighter) all the curves corresponding to all parameter vectors in the sample.

If the quantity \( \gamma(t) \) is an observable and we know approximately the observation error, we may also calculate the predictive uncertainty of new observations. Assuming that the new observations \( g(t) \) are modelled with independent additive noise,

\[
g(t) = \gamma(t) + \varepsilon(t), \quad \varepsilon \sim \pi_{\text{noise}},
\]

then the posterior predicted distribution of \( g(t) \) can be simulated by generating a sample

\[
\{ g_1(t) = \gamma_1(t) + \varepsilon_1(t), \ldots, \gamma_N(t) + \varepsilon_n(t) \}, \quad \varepsilon_j(t) \sim \pi_{\text{noise}}(\varepsilon(t)),
\]

where the predicted noise uncertainty is added to the model uncertainty.

7. Sensitivity analysis: over time and across a sample

In the previous section we have seen that the model uncertainty gives an indication of how strongly the model predictions depend on variations of the parameters according to their posterior importance. There are however reasons why the more traditional sensitivity analysis, concerned with the variability of the state of the system as a function of perturbations in the parameters, is also important. We define the sensitivity functions with respect to the parameters of the system as the logarithmic derivatives

\[
S_{j,k} = \frac{\partial \log C_j(t, \Theta, C_0)}{\partial \log \Theta_k} = \frac{\Theta_k C_j(t, \Theta, C_0)}{C_j(t, \Theta, C_0)} \frac{\partial C_j(t, \Theta, C_0)}{\partial \Theta_k},
\]
and the sensitivities with respect to the initial concentrations analogously. Sensitivity analysis is related to parameter estimation. In fact, when the number of parameters is high and model reduction by fixing some of the model parameters helps to lower the complexity of the system, it is important not to fix those parameters that have a strong influence on the state of the system of particular interest. Sensitivity analysis can help classifying the parameters in terms of their influence on particular states of the system, as proposed in [2]. Identifying less critical parameters is also important for Markov Chain Monte Carlo sampling methods, where to assess the convergence of a chain, the sample histories of individual components of the parameter vector are typically examined. In this process, it is convenient to separate influential components, which need to be well converged, from those that have only a marginal effect on the dynamic state, whose convergence is less crucial.

It is rather obvious that when the concentrations change over time, the sensitivities are also time dependent. Therefore a concentration may switch from being highly sensitive to perturbation of a parameter in a time interval, to being nearly insensitive in another. Accounting for the dynamic nature of the sensitivity functions implies that the derivatives of the concentrations with respect to the parameters which identify the dynamic mass balance equations need to be propagated in time, thus requiring the solution of a system of differential equations of dimensions equal to the product of the number of concentrations of interest times the number of parameters.

The sensitivity functions are themselves functions of the parameters. Therefore, to be consistent with the Bayesian framework, the sensitivity functions should be evaluated over a sample of parameters. This sample can be drawn from the prior if the sensitivity analysis is done before the posterior sampling, or from the posterior sample for future use: in fact, the Bayesian formulation is a learning framework, where current posterior density can be used for future prior.

The computation of the sensitivity functions is technically straightforward, albeit time consuming. Given a point $\Theta \in \mathbb{R}^K$, we define the logarithmically perturbed parameter values $\Theta^k = (\Theta_1, \ldots, \Theta_k \exp(1 + \delta)), \ldots, \Theta_K)$, where $\delta > 0$, and write the approximation

$$\frac{\partial \log C_j(t, \Theta^k, C_0)}{\partial \log \Theta_k} \approx \frac{1}{\delta} \left( \frac{C_j(t, \Theta^k, C_0)}{C_j(t, \Theta, C_0)} - 1 \right),$$

which requires the propagation of $K + 1$ state vectors over the interval of interest. This approximation gives the sensitivity at a fixed $\Theta$ in the longitudinal directions. To obtain information about the lateral variability of the sensitivity, that is, over the prior or posterior distribution of the parameter, we need to repeat the analysis at numerous sampled values $\Theta_l$ drawn from the appropriate distribution. The output flow of this procedure is tremendous, and in order to extract the desired information it is crucial to organize the throughput in an easily interpretable way. Visual modalities for the output analysis have been proposed in [2]. In Figure 6 we show the pointwise credibility envelopes of one component of the sensitivity function, and the corresponding time evolutions of the histograms, based on a time propagation of a sample of sensitivity functions for the three compartment model of the myocardial metabolism (Example 2), with the time evolution corresponding to a simulation of an ischemic episode, see [1] for details. In Figure 7, a different visualization modality is shown: the intensities of the pixels of the image correspond to the values of the sensitivities of the lactate concentration in myocardial cytosol at 7 time instances (row ordering) to the Michaelis-Menten parameters $V_{max}$ and $K$ (column ordering) for two different realizations of the parameter vector $\Theta$.

### 8. Discussion and conclusions

As our understanding of cellular metabolism increases, so does the need for computational models for hypothesis testing and output prediction. The breakthrough in gene silencing
Figure 6. Pointwise credibility envelopes of one component of the sensitivity function (left), and the corresponding time evolutions of the histogram (right), based on a time propagation of a sample of sensitivity functions for the three compartment model of the heart of Example 2, with the time evolution corresponding to a simulation of an ischemic episode.

Figure 7. Sensitivity function of the cytosolic lactate concentration with respect to the Michaelis-Menten parameters $V_{\text{max}}$ and $K$. The sensitivities, calculated at two different parameter points $\Theta$, are displayed in matrix form.

and knockout research, whose visibility has recently reached the general public with the awarding of the 2007 Nobel prize in physiology and medicine has brought new understanding of the functioning of enzymes and transporters in the cells. Concurrently, new modalities for acquiring biological data at the cellular and subcellular level, via functional quantitative imaging modalities such as NMR spectroscopy, fMRI and PET, continue to improve our knowledge of cellular processes. The experimental and technological advantages have reinforced the notion that computational models are not only useful but necessary to have coherent picture and to retrieve pertinent information from indirect observations. The field of systems biology and the quest for understanding how to cure and prevent metabolic diseases promises to be a great source of challenges for applied mathematics and in particular for the field of inverse problems.
References


