# AALBORG UNIVERSITY

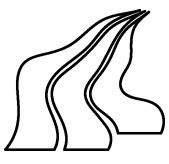
## A Population based Bayesian Approach to the Minimal Model of Glucose and Insulin Homeostasis

by

Kim E. Andersen and Malene Højbjerre

December 2003 R-2003-25

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: www.math.auc.dk/research/reports/reports.htm



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

# A Population based Bayesian Approach to the Minimal Model of Glucose and Insulin Homeostasis

Kim E. Andersen and Malene Højbjerre*

*Department of Mathematical Sciences, Aalborg University,*
*Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark*

SUMMARY

The classical minimal model of glucose and insulin homeostasis was proposed in the late seventies by Bergman et al. (1979) [1] as a powerful model consisting of three differential equations describing the metabolic portrait of a single individual, playing an important role in the study of diabetes. Considering both the glucose and insulin kinetics simultaneously, the minimal model becomes a highly ill-posed problem, and the reconstruction has therefore most often been done by non-linear least squares techniques separately for each entity. Hereby the severe ill-posedness is ignored and the estimation may only be efficient when good initial estimates are provided. Furthermore the minimal model is only specified for a single individual and does not provide the opportunity for using information from other individuals, if available, with the advantage of estimating the metabolic portrait for a whole population. Finally, the minimal model has only been analyzed in a deterministic set-up with only error terms on the measurements. In this work we adopt a Bayesian graphical model to describe a stochastic version of the coupled minimal model, meaning that we do not only consider error terms on the observations, but also on the process increments. Furthermore we describe the minimal model for a whole population, including a single individual. The estimation of the parameters are efficiently implemented in a Bayesian approach where posterior inference is made through the use of Markov chain Monte Carlo techniques. Hereby we obtain a powerful and flexible modeling framework for regularizing the ill-posed estimation problem often inherited in coupled stochastic differential equations. We demonstrate the method on experimental data from intravenous glucose tolerance tests performed on 19 normal glucose tolerant subjects.

**Keywords:** Glucose and insulin homeostasis; population minimal model; Bayesian graphical model; MCMC methods; coupled differential equations, stochastic differential equations.

## 1. INTRODUCTION

Diabetes mellitus is a metabolic disease associated with a number of abnormalities in the insulin metabolism, ranging from an absolute deficiency to a combination of deficiency and

*Correspondence to: Malene Højbjerre, Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark

email: malene@math.auc.dk

resistance, causing an inability to dispose glucose from the blood stream at normal rates. Three factors, referred to as the metabolic portrait of a single individual [2], play an important role for glucose disposal: (1) *insulin sensitivity*, specifying the insulin's capability to increase the glucose disposal to muscles, liver and adipose tissue; (2) *glucose effectiveness*, representing the ability of the glucose in plasma to enhance its own disposal at basal insulin level; and (3) *pancreatic responsiveness*, characterizing the ability of the pancreatic $\beta$-cells to secrete insulin in response to glucose stimuli. Failure in any of these may lead to impaired glucose tolerance, or, if severe, diabetes. Quantitative assessment of these factors is made possible by the the minimal model [1], and may improve classification, prognosis and therapy of diabetes [3].

The minimal model is based on data from an intravenous glucose tolerance test (IVGTT), where glucose and insulin concentrations in plasma are subsequently sampled after an intravenous glucose injection. In the minimal model the glucose and insulin kinetics are separately described by two sets of differential equations, and the parameters have traditionally been estimated by a non-linear weighted least squares estimation technique separately within each set of differential equations, as described by Pacini and Bergman (1986) [2] for example. There are many problems associated with this approach, for example negative confidence intervals for strictly positive parameters and insulin sensitivity predicted equal to zero. Pillonetto et al. (2002) [4] adopt a Bayesian approach to this problem treating the insulin as a known forcing function, i.e. the model is only analyzed for glucose kinetics. The basis for the minimal model assumes that the glucose and insulin constitute a single dynamical system and important information is lost in trying to treat the system in two separate parts. Nevertheless, as pointed out by de Gaetano and Arino (2000) [5] coupling of the glucose and insulin leads to a highly ill-posed problem, where adequate reconstruction of the system depends on proper regularization and good initial values.

The minimal model was originally defined for a single individual providing only estimation of separate individual metabolic portraits. Traditionally there has been no sharing of information between individuals when considering a whole population. A population based approach with the aim of estimating the metabolic portrait for a whole population, e.g. the normal population, can be very useful in the study of diabetes for classification of patients being in a risk group or not of developing diabetes. Furthermore, information about the population can be utilized as prior information in the statistical analysis for a single individual. Both Vicini and Cobelli (2001) [6] and Agbaje et al. (2003) [7] propose a Bayesian approach to the population based minimal model, but again only the glucose part of the model is considered.

Earlier in the literature, including the above-mentioned papers, the minimal model has solely been considered as a deterministic model with only error terms on the observations. We believe that the differential equations describing the glucose and insulin processes may not comply with the actual processes taking place inside the body. Therefore we propose a stochastic version of the minimal model, where error terms on the process increments are included, besides error terms on the observations. Furthermore, we combine both the glucose and the insulin parts of the minimal model to obtain a unified model, which is more physiological sound but also more complex and computationally intractable. However, by adopting a Bayesian graphical model [8] we are able to describe the stochastic minimal model first for a single individual, and then for a whole population. In order to assess the parameters we regularize the ill-posed problem by use of use prior knowledge. This is done in a Bayesian approach, where posterior inference is made by use of Markov chain Monte Carlo (MCMC) methods [9].

In Section 2 we discuss the minimal model for glucose and insulin kinetics for a single individual. In Section 3 we introduce Bayesian graphical model and the computational techniques required for inference in this model. The minimal model for a single individual and for a whole population are expressed as Bayesian graphical models in Section 4 and Section 5, respectively. Specific implementational issues within the simulations algorithm for performing inference in these models are addressed in Section 6. In section 7 we consider the behavior and performance of our approach on field data and finally in Section 8 we discuss our method and further potential developments.

## 2. THE MINIMAL MODEL OF GLUCOSE-INSULIN HOMEOSTASIS

In a controlled standard frequently sampled IVGTT study a small glucose load is administered intravenously to overnight-fasted subjects for the purpose of recording the responding glucose and insulin concentrations in plasma. Approximately 20 to 30 minutes before glucose is administered, two cannulae are inserted into an antecubical vein in both arms and the patency of the cannulae is maintained with a controlled saline infusion throughout the complete study. Typically three blood samples are obtained prior to the glucose injection, e.g. immediately before the injection and at $-15$ and $-30$ minutes, for determination of the basal insulin and glucose concentrations in plasma. At time 0, a dose of glucose is injected (usually 0.3 gram glucose per kilo body weight) over a 30 – 60 seconds period into the cannula contralateral to the one used for blood sampling and subsequently several (usually 25 – 30) venous blood samples (3 – 10 ml) are acquired over a period of typically 180 to 240 minutes. The obtained blood samples are immediately centrifuged under refrigeration and analyzed for glucose and insulin concentration in plasma. Glucose and insulin time courses on a log-scale from a frequently sampled IVGTT for a normal glucose tolerant individual are depicted in Figure 1. In this particular case 23 blood samples are taken at $-30$, $-15$, 0, 2, 4, 6, 8, 10, 15, 20, 25, 30, 40, 50, 60, 75, 90, 105, 120, 150, 180, 210 and 240 minutes.

It is apparent from Figure 1, that the injected glucose load immediately elevates the glucose concentration in plasma initiating secretion of insulin from the pancreatic $\beta$-cells. The provoked hyperglycemia induces an immediate peak in the insulin concentration in plasma, and the glucose uptake in muscles, liver and adipose tissue is raised by the actual effect of the insulin, the so-called remote insulin in action. This lowers the glucose concentration in plasma, affecting the $\beta$-cells to secrete less insulin, from which a feedback effect arises. By approximately 60 minutes, the glucose concentration is normalized, and in the following two hours a moderate undershoot is observed. After approximately 2 – 3 hours, it is usually found that the perturbed insulin and glucose concentrations essentially have returned to normal. Often a second phase pancreatic responsiveness follows at the immediate peak in insulin induced by hyperglycemia. However, depending upon the state of the tested subject, the glucose and insulin time courses may vary considerably from the response shown in Figure 1.

The minimal model, proposed by Bergman et al. (1979) [1] and further developed in Bergman et al. (1981) [10], describes this dynamic system of both the glucose and insulin kinetics. The minimal model is represented by a compartmental system, as illustrated in Figure 2. The plasma insulin space is represented as a single extra-cellular compartment from which plasma insulin at time $t$, denoted by $I(t)$, is believed to fill a remote active intra-cellular

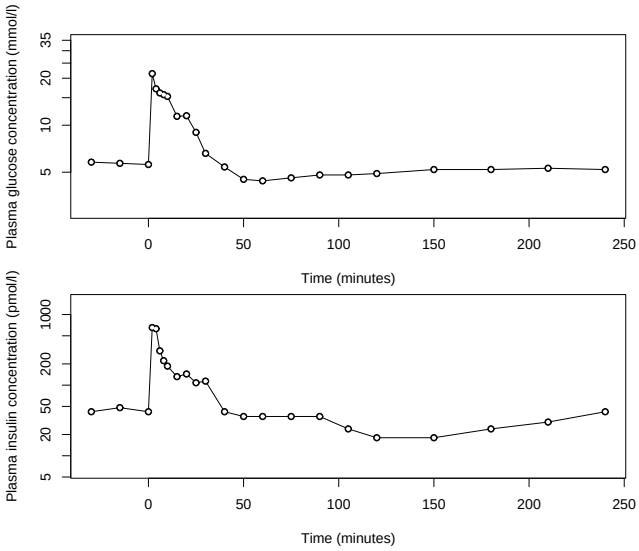



Figure 1. Glucose and insulin concentrations in plasma frequently sampled over 240 minutes after an intravenous glucose injection given to a normal glucose tolerant individual.

insulin compartment with a constant rate $p_3$. The active insulin in the remote compartment is presumably bound insulin which is active in accelerating glucose utilization by peripheral tissues and the liver. Consequently $p_3$ is referred to as the insulin-dependent increase in glucose uptake ability in tissue. The intra-cellular metabolism of the remote insulin effect is constant and determined by $p_2$, i.e. $p_2$ is the rate expressing the impulsive decrease of tissue glucose uptake ability. The dynamic insulin response, denoted by $X(t)$, is proportional to the active insulin in the remote compartment and describes the time dependent effect of the insulin on the net glucose disappearance. The plasma glucose at time $t$, denoted by $G(t)$, is represented by a single extra-cellular compartment and the glucose concentration in plasma is governed by a balance between the glucose production/uptake by the liver and the utilization of glucose by the peripheral tissues. This insulin-independent glucose uptake in muscles, liver and tissue is assumed to be of constant rate $p_1$.

Consequently the equations of the minimal model for glucose disposal are

$$\begin{aligned} \dot{G}(t) &= -p_1(G(t) - G_b) - X(t)G(t), \qquad & G(0) &= G_0, \\ \dot{X}(t) &= -p_2 X(t) + p_3(I(t) - I_b), & X(0) &= 0, \end{aligned} \tag{1}$$

where $G_b$ and $I_b$ denote basal preinjection levels of glucose and insulin, respectively, and $G_0$ is the theoretical glucose concentration in plasma extrapolated to the time of glucose injection, i.e. at time $t = 0$.

The insulin secretion is described by two phases: (1) a first-phase insulin secretion of constant rate $n$ independent of glucose stimuli; and (2) a second-phase insulin secretion proportional to the degree by which glucose exceeds a certain threshold level $h$, envisioning that the $\beta$-cells only release insulin whenever glucose exceeds $h$. The minimal model of insulin kinetics is then

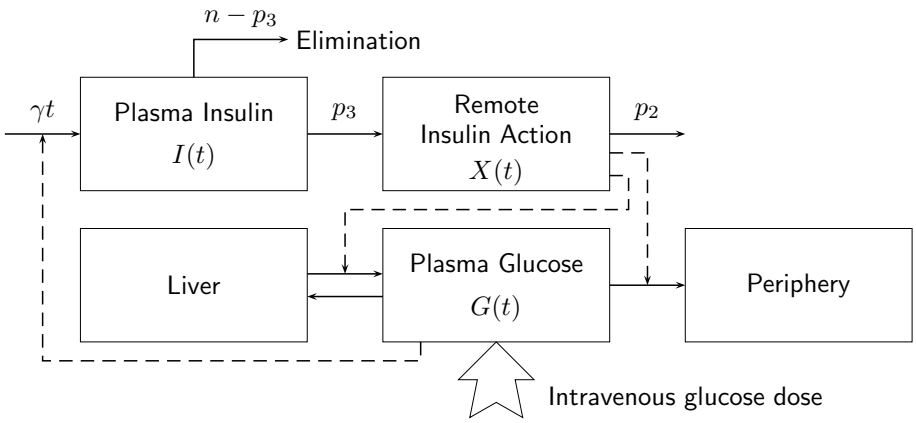



Figure 2. The minimal model for a single individual describing the glucose and insulin kinetics in an IVGTT study.

given by the following non-linear differential equation

$$\dot{I}(t) = -n(I(t) - I_b) + \gamma(G(t) - h)t, \qquad I(0) = I_0, \tag{2}$$

where $I_0$ denotes the theoretical insulin concentration in plasma extrapolated to time $t = 0$ and $\gamma$ denotes the rate by which the $\beta$-cells release insulin. The multiplication of time $t$ in the second-phase insulin secretion was introduced in order to formulate the assumption that the pancreatic release of insulin is not only proportional to the hyperglycemia attained but also to the time elapsed since glucose injection [11].

The parameter $p_1$ describes the rate of net glucose utilization if there were no dynamic insulin response, i.e. $X(t) \equiv 0$. Thus $p_1$ is a measure of the ability of glucose to enhance its own disappearance within the extra-cellular compartment at basal insulin level independent of any increment in the plasma insulin. In the literature, usually $p_1$ is referred to as the *glucose effectiveness*, denoted by $S_G = p_1$. Of interest is also the *insulin sensitivity* defined by the fraction $S_I = p_3/p_2$, as this represents the capability of insulin to increase the net glucose utilization. The two parameters $S_G$ and $S_I$ play an important role for the characterization of an individuals glucose disposal. In addition, however, it is also possible from the model parameters in (1) and (2) to obtain a quantitative description of the pancreatic sensitivity to any change in plasma glucose concentration. As discussed above, the insulin secretion and its sensitivity to glucose stimuli is described by a first- and second-phase *pancreatic responsiveness*, $\varphi_1$ and $\varphi_2$, respectively. These are defined as

$$\varphi_1 = \frac{I_0 - I_b}{n(G_0 - G_b)},$$
$$\varphi_2 = \gamma \times 10^4,$$

see e.g. Toffolo et al. (1980) [11] for details.

Essentially all strategies employed in the literature for assessment of the metabolic portrait, see e.g. Pacini and Bergman (1986) [2], resort to an iterative non-linear weighted least squares

estimation technique, where the insulin is treated as known, i.e. the minimal model is regarded as a deterministic model with only measurement errors and composed of only the component defined in (1). Hereby the model identification strategies are purely data-driven and do not consider the integrated glycemic system formed by the pancreas and tissues with the proper feedback regulations that exists simultaneously. A frequently encountered problem with this iterative non-linear weighted least squares approach is that, especially for subjects diagnosed with type II diabetes, it provides a point estimate of insulin sensitivity equal to zero or close to zero [12; 13; 14]. Moreover, since the corresponding confidence intervals are based upon the Fisher information matrix, negative values may easily be included. Furthermore the MINMOD programme described in Pacini and Bergman (1986) [2] only estimates $S_I$ and $S_G$ and not $\varphi_1$ and $\varphi_2$. A more advanced estimation technique that allows for physically sound parameter estimation while considering both (1) and (2) simultaneously are therefore called for. Besides, it seems reasonable to consider the minimal model as a stochastic differential equation model due to expected fluctuations from the underlying processes.

However, solving the three coupled differential equations in (1) and (2) simultaneously (deterministic or stochastic) is a highly ill-posed problem, and it can easily be shown, that for even commonly observed combinations of parameter values the system may not admit a well-defined equilibrium, see e.g. de Gaetano and Arino (2000) [5] for details. Consequently regularization of the coupled minimal model is needed. Ill-posed problems are often regularized by imposing certain regularity conditions on the solution space. This is equivalent to using a penalized likelihood, where the solution space is reduced by introducing a penalty function for implausible solutions, or in the Bayesian approach using a prior, where the implausible parameters automatically are penalized. There are other advantages inherent in the Bayesian approach, e.g. the availability of computational tools (such as MCMC methods) which allow the construction and analysis of suitably complex models without the need for simplifying assumptions.

In our approach we adopt a complex graphical model to represent the quite complicated relationship among the quantities of the stochastic minimal model, and analyze this model in a Bayesian approach, whereby the coupled minimal model is regularized. A further advantage of this approach is a modeling framework of hierarchical models, in which the minimal model very easily can be extended to a population based model where information can be shared between individuals meaning less informative priors and a smaller number of observations and simulations are expected to be required for appropriate inference. In the next section we review Bayesian graphical models, and discuss MCMC methods which facilitate Bayesian inference in these models.

## 3. GENERAL ASPECTS OF THE STATISTICAL MODELING

A directed graphical model represents complicated dependencies among quantities of a model by introducing appropriate conditional independence assumptions, which conveniently can be represented by a graph, where the vertices represent quantities of the model and the

missing links represent conditional independence assumptions. In this section we give a brief introduction to directed graphical models. For a more detailed discussion on graphical models we refer to Lauritzen (1996) [8]. We also consider how to perform Bayesian inference in directed graphical models by the use of MCMC methods, which over the past decade have enjoyed widespread popularity within the statistical literature. For a general introduction to this field we suggest Robert and Casella (1999) [9], for example.

### *3.1. Directed Graphical Models*

Let $\mathcal{G} = (V, E)$ be a directed acyclic graph (DAG), where $V$ is a finite set of vertices, corresponding to all the quantities of the model, and $E$ is a finite set of directed edges, illustrating a direct influence from one quantity to another. The set $E$ is a subset of ordered pairs of distinct vertices in $V$, and the graph being acyclic means that by following the directions of the edges, it is not possible to return to a vertex visited before. In notation we do not distinguish between a vertex and its corresponding random variable.

A directed graphical model is defined by the joint distribution of all the vertices, $V$, admitting the directed Markov property w.r.t. $\mathcal{G}$. There are several equivalent ways to state the directed Markov property – one is to say that the joint density of all the vertices, $p(V)$, admits the recursive factorization property given as

$$p(V) = \prod_{v \in V} p(v \mid \mathrm{pa}(v)), \tag{3}$$

where $p(v|\,\mathrm{pa}(v))$ is the conditional density of the vertex $v$ given its parents $\mathrm{pa}(v)$. This means that the joint density of all the vertices factorizes into the product of all the parent-child densities, whereby the model is totally specified by these local conditional densities. Furthermore, it will be apparent that the local parent-child distributions are also all needed to perform Bayesian computations in directed graphical models by MCMC methods.

The directed Markov property implies that the graph can be used to read off conditional independence assumptions for the model. For example, the directed local Markov property states that each vertex $v$ is conditional independent of all its non-descendants given its parents, which for all $v \in V$ is written as

$$v \perp\!\!\!\perp \mathrm{nd}(v) \,|\, \mathrm{pa}(v).$$

### *3.2. Bayesian Inference*

In Bayesian inference all quantities of the statistical model are considered as random variables. To distinguish between the different kinds of quantities we let $\Phi$ denote the observed data and $\Theta$ the model parameters. Conceptually the parameters may also include other unobserved quantities of the model, such as unobserved data and/or latent variables. The main focus in Bayesian inference is the posterior distribution of the unobserved random variables given the observed random variables, $p(\Theta \,|\, \Phi)$, as it represents our beliefs about the feasible structures of $\Theta$ after having observed the data $\Phi$. To achieve this posterior, a prior distribution representing our beliefs about the parameters before having observed any data, $p(\Theta)$, is required. The posterior distribution is then determined by

$$p(\Theta \,|\, \Phi) \propto L(\Theta \,|\, \Phi) p(\Theta)$$

which expresses the posterior as the likelihood, $L(\Theta \mid \Phi)$, i.e. the information about the parameters from data, modified by our prior knowledge about the parameters, the prior distribution $p(\Theta)$.

Performing inference about the model parameters are hereby reduced to the computational task of calculating posterior moments, which means evaluating integrals over the parameter space $\boldsymbol{\Theta}$, for example the posterior mean

$$\mathbb{E}(\Theta) = \int_{\boldsymbol{\Theta}} \Theta \, p(\Theta \mid \Phi) \, d\Theta.$$

Explicit evaluation of such integrals are often analytically intractable or even impossible due to the huge state space, however, MCMC methods provide an approximative integration technique whereby marginal posterior means, for example, are estimated by the sample mean of a representative series of random draws from the posterior distribution. These random draws are obtained by constructing an irreducible Markov chain $\Theta_1, \Theta_2, \ldots$ with state space $\boldsymbol{\Theta}$ and with stationary distribution $p(\Theta \mid \Phi)$. MCMC sampling was first introduced by Metropolis et al. (1953) [15] and was subsequently adapted by Hastings (1970) [16]. There exist various standard techniques for constructing such Markov chains, see e.g. Brooks (1998) [17] or Robert and Casella (1999) [9]. In this paper we will utilize the Metropolis–Hastings updates.

### *3.3. Metropolis–Hastings Updates in Bayesian Graphical Models*

Metropolis–Hastings updates are used to move around the parameter space by proposing moves which are subsequently either accepted or rejected. Suppose that we are currently in state $\Theta$, and then we draw a new state $\Theta'$ from some proposal density $q(\Theta; \Theta')$. This proposal is then accepted with probability

$$\alpha(\Theta; \Theta') = \min \left\{ 1, \frac{p(\Theta' \mid \Phi) q(\Theta'; \Theta)}{p(\Theta \mid \Phi) q(\Theta; \Theta')} \right\}.$$

However, if the proposal is rejected, the chain remains in the current state. Many proposal distributions lead to irreducible Markov chains which ensure the convergence of the posterior mean estimate, though several forms possess useful analytic properties. For example, when the proposal distribution $q$ is symmetric, i.e. $q(\Theta; \Theta') = q(\Theta'; \Theta)$, the acceptance probability reduces to

$$\alpha(\Theta; \Theta') = \min \left\{ 1, \frac{p(\Theta' \mid \Phi)}{p(\Theta \mid \Phi)} \right\},$$

which is essentially the original Metropolis update proposed by Metropolis et al. (1953) [15].

In directed graphical models a simple way to update $\Theta$ is to successively update each unobserved vertex $v$ one after another. Being in the current state $v$ and proposing a new state $v'$ from a proposal distribution $q(v; v')$, the acceptance probability hereby becomes

$$\alpha(v; v') = \min \left\{ 1, \frac{p(v' \mid \operatorname{pa}(v)) \prod_{w : v' \in \operatorname{pa}(w)} p(w \mid \operatorname{pa}(w)) q(v'; v)}{p(v \mid \operatorname{pa}(v)) \prod_{w : v \in \operatorname{pa}(w)} p(w \mid \operatorname{pa}(w)) q(v; v')} \right\},$$

by using the recursive factorization in (3). This implies that it is only the parent-child densities of the corresponding Markov blanket (the vertex itself, its parents, its children and its children's

other parents) that are necessary in each update of a single vertex. These successive updates of all the unobserved vertices ensures a construction of a Markov chain with stationary distribution $p(\Theta \,|\, \Phi)$, and the sampling in a huge parameters space is reduced to only local computations in small subsets of the graph. Concerning the proposal distribution $q(v; v')$ it can be advantageous to use the conditional distribution of a vertex given the rest of the vertices if this distribution is of standard form. This distribution is in fact proportional to

$$P(v \,|\, V \setminus v) \propto p(v \mid \mathrm{pa}(v)) \prod_{w : v \in \mathrm{pa}(w)} p(w \mid \mathrm{pa}(w))$$

whereby the acceptance probability reduces to 1, and every proposal is accepted. This special updating scheme is also known as the Gibbs sampler, see e.g. [18]. The use of MCMC methods in Bayesian graphical models has been treated in details in Spiegelhalter (1998) [19] and Lunn et al. (2000) [20].

In the following section we adopt a directed graphical model to represent a stochastic version of the minimal model given by both (1) and (2).

## 4. THE MINIMAL MODEL FOR A SINGLE INDIVIDUAL AS A BAYESIAN GRAPHICAL MODEL

The glucose and insulin concentrations are positive, and experience shows that the variability in the blood samples increases with the mean. Therefore we assume that both $G(t)$, $X(t)$ and $I(t)$ are log-normally distributed and introduce $g(t) = \log G(t)$, $x(t) = \log X(t)$ and $i(t) = \log I(t)$, where $g(t)$, $x(t)$ and $i(t)$ are normally distributed. This logarithmic transformation implies that $\dot{g}(t) = \dot{G}(t)/G(t)$, $\dot{x}(t) = \dot{X}(t)/X(t)$ and $\dot{i}(t) = \dot{I}(t)/I(t)$. The minimal model for a single individual given by (1) and (2) can then be rewritten as

$$\begin{aligned}
\dot{g}(t) &= -S_G(1 - G_b\, e^{-g(t)}) - e^{x(t)}, & g(0) &= \log G_0, \\
\dot{x}(t) &= -p_2(1 - S_I(e^{i(t)} - I_b)e^{-x(t)}), & x(0) &\to -\infty, \\
\dot{i}(t) &= -\frac{I_0 - I_b}{\varphi_1(G_0 - G_b)}(1 - e^{-i(t)} I_b) + 10^{-4} e^{-i(t)} \varphi_2 (e^{g(t)} - h)t, & i(0) &= \log I_0,
\end{aligned} \tag{4}$$

where we have re-parameterized the model according to the metabolic portrait given by the four parameters $S_G$, $S_I$, $\varphi_1$ and $\varphi_2$. The logarithmic transformed minimal model has no unit of measurement, and the three processes can be re-parameterized, such that they are on the same scale.

The glucose and insulin system described by the deterministic minimal model in (4) may, however, not comply with the actual log-transformed glucose and insulin time courses from an IVGTT study. We therefore introduce a stochastic version of the minimal model, where Brownian motion fluctuations $B^g$, $B^x$ and $B^i$ are used to model possible model deviations, i.e. the stochastic minimal model takes the differential form

$$\begin{aligned}
dg(t) &= \big( - S_G(1 - G_b e^{-g(t)}) - e^{x(t)} \big) dt + \tau_g^{-1/2} dB^g(t), \\
dx(t) &= \big( - p_2(1 - S_I(e^{i(t)} - I_b)e^{-x(t)}) \big) dt + \tau_x^{-1/2} dB^x(t), \\
di(t) &= \Big( - \frac{I_0 - I_b}{\varphi_1(G_0 - G_b)}(1 - e^{-i(t)} I_b) + 10^{-4} e^{-i(t)} \varphi_2 (e^{g(t)} - h)t \Big) dt + \tau_i^{-1/2} dB^i(t),
\end{aligned}$$

where $\tau_g$, $\tau_x$ and $\tau_i$ denote the reciprocal variances (the so-called precisions) of the Brownian motions. The analysis of the differential form of the stochastic minimal model can be transferred by simple integration to that of an equivalent set of integral equations, e.g. for $g(t)$ we obtain

$$g(t_k) - g(t_{k-1}) = \int_{t_{k-1}}^{t_k} (-S_G(1 - G_b e^{-g(t)}) - e^{x(t)})dt + \tau_g^{-1/2}(B^g(t_k) - B^g(t_{k-1})),$$

where $t_k - t_{k-1} > 0$ is a suitably small time span. The involved unknown integral is approximated by the product between its width and its left end point, that is

$$g(t_k) = g(t_{k-1}) - (t_k - t_{k-1})\big(S_G(1 - G_b e^{-g(t_{k-1})}) + e^{x(t_{k-1})}\big) + \epsilon^g(t_k - t_{k-1}),$$

where the random process $\epsilon^g(t_k - t_{k-1}) = \tau_g^{-1/2}(B^g(t_k) - B^g(t_{k-1}))$ is well-known to depend only on the time interval $t_k - t_{k-1}$ and to follow a normal distribution with mean zero and variance $\tau_g^{-1}(t_k - t_{k-1})$. Note, adding noise on the process increments corresponds to assuming that the latent processes are Itô processes [21].

If we introduce a more convenient notation using $t$ as a subscript, e.g. $g(t_k) = g_{t_k}$, then the stochastic minimal model can be rewritten as

$$\begin{aligned} g_{t_k} &= f^g(g_{t_{k-1}}, x_{t_{k-1}}) + \epsilon^g(t_k - t_{k-1}), \\ x_{t_k} &= f^x(x_{t_{k-1}}, i_{t_{k-1}}) + \epsilon^x(t_k - t_{k-1}), \\ i_{t_k} &= f^i(i_{t_{k-1}}, g_{t_{k-1}}) + \epsilon^i(t_k - t_{k-1}), \end{aligned}$$

with

$$\begin{aligned} f^g(g_{t_{k-1}}, x_{t_{k-1}}) &= g_{t_{k-1}} - (t_k - t_{k-1})\big(S_G(1 - G_b\, e^{-g_{t_{k-1}}}) + e^{x_{t_{k-1}}}\big), \\ f^x(x_{t_{k-1}}, i_{t_{k-1}}) &= x_{t_{k-1}} - (t_k - t_{k-1})p_2\big(1 - S_I(e^{i_{t_{k-1}}} - I_b)e^{-x_{t_{k-1}}}\big), \\ f^i(i_{t_{k-1}}, g_{t_{k-1}}) &= i_{t_{k-1}} + (t_k - t_{k-1})\Big(-\frac{I_0 - I_b}{\varphi_1(G_0 - G_b)}(1 - e^{-i_{t_{k-1}}} I_b) + 10^{-4} e^{-i_{t_{k-1}}} \varphi_2(e^{g_{t_{k-1}}} - h)t_{k-1}\Big), \end{aligned}$$

where we for notational convenience have suppressed the functional dependencies of the parameters $S_G, S_I, \varphi_1, \varphi_2, p_2, h, G_b, I_b, G_0, I_0$ and the time.

The conditional distributions for the processes $g_{t_k}$, $x_{t_k}$ and $i_{t_k}$ are hereby given as

$$\begin{aligned} g_{t_k} \,|\, g_{t_{k-1}}, x_{t_{k-1}}, \tau_g &\sim \mathcal{N}(f^g(g_{t_{k-1}}, x_{t_{k-1}}), \tau_g^{-1}(t_k - t_{k-1})), \\ x_{t_k} \,|\, x_{t_{k-1}}, i_{t_{k-1}}, \tau_x &\sim \mathcal{N}(f^x(x_{t_{k-1}}, i_{t_{k-1}}), \tau_x^{-1}(t_k - t_{k-1})), \\ i_{t_k} \,|\, i_{t_{k-1}}, g_{t_{k-1}}, \tau_i &\sim \mathcal{N}(f^i(i_{t_{k-1}}, g_{t_{k-1}}), \tau_i^{-1}(t_k - t_{k-1})). \end{aligned} \tag{5}$$

These conditional distributions can be interpreted as parent-child distributions in a directed graphical model, that can be illustrated by the directed acyclic graph in Figure 3.

For the clearness of the graph we have omitted the parameter vertices $S_G, S_I, \varphi_1, \varphi_2, p_2, h, G_b, I_b, G_0, I_0, \tau_g, \tau_i$ and $\tau_x$, but added the random variables $g^o_{t_k} = \log(G^o_{t_k})$ and $i^o_{t_k} = \log(I^o_{t_k})$, where $G^o_{t_k}$ and $I^o_{t_k}$ are the random variables actually observed for specific time points $t_k$. To distinguish between the different types of vertices we illustrate the observed vertices by rectangles and the unobserved vertices by circles.

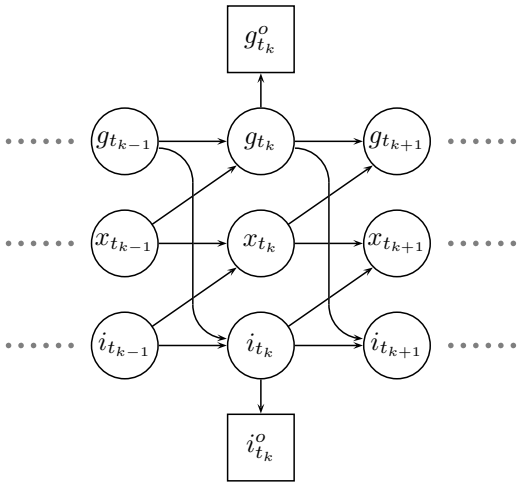



Figure 3. Directed acyclic graph illustrating the statistical dependencies for the latent processes, $g$, $x$ and $i$, and the observed processes $i^o$ and $g^o$. The observed quantities are illustrated by rectangles and the unobserved quantities are illustrated by circles.

We model the measurement error on $g^o_t$ and $i^o_t$ by two independent random white noise processes; $\epsilon^{g^o}$ and $\epsilon^{i^o}$ with precisions $\tau_{g^o}$ and $\tau_{i^o}$, respectively. Consequently the distributional assumptions for $g^o_{t_k}$ and $i^o_{t_k}$ are

$$\begin{aligned} g^o_{t_k} \mid g_{t_k}, \tau_{g^o} &\sim \mathcal{N}(g_{t_k}, \tau^{-1}_{g^o}), \\ i^o_{t_k} \mid i_{t_k}, \tau_{i^o} &\sim \mathcal{N}(i_{t_k}, \tau^{-1}_{i^o}). \end{aligned} \tag{6}$$

Notice that the mean structures of the logarithmically transformed observations depend on the underlying non-observable and hereby latent system processes $g_{t_k}, x_{t_k}$ and $i_{t_k}$.

Let the unobserved quantities of the model, $\Theta$, be divided into two disjoint subsets consisting of the parameters

$$\Omega = (S_G, S_I, \varphi_1, \varphi_2, p_2, h, G_b, I_b, G_0, I_0, \tau_g, \tau_x, \tau_i, \tau_{g^o}, \tau_{i^o})$$

and the three latent processes

$$\Psi = \{g_{t_k}, x_{t_k}, i_{t_k}\}_{t_k \in \Lambda},$$

where $\Lambda$ denotes the time points chosen for approximation of the latent processes. Thus $\Theta = (\Omega, \Psi)$. Furthermore we let the logarithmically transformed observations be denoted by $\Phi = \{g^o_{t_k}, i^o_{t_k}\}_{t_k \in \mathcal{T}}$, where $\mathcal{T} \subseteq \Lambda$ denotes the set of actual observation times. Dividing all the quantities into these three subsets, the statistical dependencies in the model defined by (5) and (6) can be summarized by the simple directed acyclic graph in Figure 4, where the time aspect eventually has vanished. This blocking of the unobserved variables implies that the posterior distribution factorizes as

$$p(\Omega, \Psi \mid \Phi) \propto p(\Phi \mid \Omega, \Psi) p(\Psi \mid \Omega) p(\Omega),$$

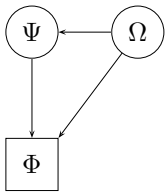



Figure 4. Directed acyclic graph illustrating the statistical relationship for the observed processes, $\Phi$, the latent processes, $\Psi$, and the parameters, $\Omega$.

where $p(\Omega)$ represents our beliefs about the parameters before having observed any data and $p(\Psi \,|\, \Omega)$ and $p(\Phi \,|\, \Omega, \Psi)$ are determined by (5) and (6) and form the likelihood.

Using the recursive factorization of the directed graphical model in Figure 3 it is easily shown that

$$\begin{aligned} p(\Psi \,|\, \Omega) &= \prod_{t_k \in \Lambda} p(g_{t_k} \,|\, g_{t_{k-1}}, x_{t_{k-1}}, \tau_g) p(x_{t_k} \,|\, x_{t_{k-1}}, i_{t_{k-1}}, \tau_x) p(i_{t_k} \,|\, i_{t_{k-1}}, g_{t_{k-1}}, \tau_i) \\ &\propto (\tau_g \tau_x \tau_i)^{N/2} \exp\{-V(\Psi, \Omega)\}, \end{aligned} \tag{7}$$

where $N = |\Lambda|$ denotes the number of elements in $\Lambda$ and the posterior potential $V$ is given by

$$V(\Psi, \Omega) = \frac{1}{2} \sum_{t_k \in \Lambda} \left\{ \tau_g (g_{t_k} - f^g_{t_k})^2 + \tau_x (x_{t_k} - f^x_{t_k})^2 + \tau_i (i_{t_k} - f^i_{t_k})^2 \right\},$$

with $f^g_{t_k} = f^g(g_{t_{k-1}}, x_{t_{k-1}})$, $f^x_{t_k} = f^x(x_{t_{k-1}}, i_{t_{k-1}})$ and $f^i_{t_k} = f^i(i_{t_{k-1}}, g_{t_{k-1}})$.

Similarly, by application of the recursive factorization property of the directed graphical model, it can be shown that

$$\begin{aligned} p(\Phi \,|\, \Omega, \Psi) &= \prod_{t_k \in \mathcal{T}} p(g^o_{t_k} \,|\, g_{t_k}, \tau_{g^o}) p(i^o_{t_k} \,|\, i_{t_k}, \tau_{i^o}) \\ &\propto (\tau_{g^o} \tau_{i^o})^{M/2} \exp\{-W(\Phi, \Omega, \Psi)\}, \end{aligned} \tag{8}$$

where $M = |\mathcal{T}|$ denotes the number of observations and

$$W(\Phi, \Omega, \Psi) = \frac{1}{2} \sum_{t_k \in \mathcal{T}} \tau_{g^o} (g^o_{t_k} - g_{t_k})^2 + \tau_{i^o} (i^o_{t_k} - i_{t_k})^2.$$

Concerning $\Omega$ we assume that the elements are independent a priori and that each of the positive system parameters $S_G, S_I, \varphi_1, \varphi_2, p_2, h, G_b, I_b, G_0$ and $I_0$ are log-normally distributed and that the positive precisions $\tau_g, \tau_x, \tau_i$, $\tau_{g^o}$ and $\tau_{i^o}$ each has a uniform shrinkage prior. Consequently $p(\Omega)$ takes the simple form

$$p(\Omega) = p(S_G)p(S_I)p(\varphi_1)p(\varphi_2)p(p_2)p(h)p(G_b)p(I_b)p(G_0)p(I_0)p(\tau_g)p(\tau_x)p(\tau_i)p(\tau_{g^o})p(\tau_{i^o}), \tag{9}$$

where the densities of the system parameters are log-normal whereas uniform shrinkage priors are used for all the precisions.

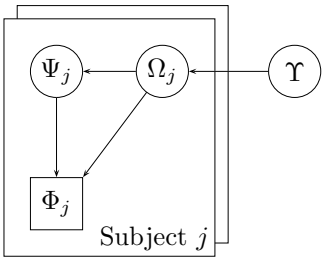



Figure 5. Directed acyclic graph illustrating the statistical relationship in the population based minimal model, where individual $j$ has the observed processes, $\Phi_j$, the latent processes, $\Psi_j$, and the parameters, $\Omega_j$ sampled from the population distribution with parameter $\Upsilon$.

## 5. THE MINIMAL MODEL FOR A POPULATION AS A BAYESIAN GRAPHICAL MODEL

Having specified the minimal model for a single individual as a directed graphical model, it can easily be extended to a hierarchical population based model.

Suppose we consider $L$ independent subjects, where subject $j$, for $j = 1, \ldots, L$, has its own individual observed processes, $\Phi_j$, own latent processes, $\Psi_j$, and own parameters, $\Omega_j$. We assume that the individual system parameters, $S_{Gj}, S_{Ij}, \varphi_{1j}, \varphi_{2j}, p_{2j}, h_j, G_{bj}, I_{bj}, G_{0j}$ and $I_{0j}$, including the individual metabolic portrait, will arise from a population distribution with parameters $\Upsilon$, where $\Upsilon$ consists of independent population means and population precisions for the log-normal distributed parameters, i.e.

$$\Upsilon = (\mu_{S_G}, \tau_{S_G}, \mu_{S_I}, \tau_{S_I}, \mu_{\varphi_1}, \tau_{\varphi_1}, \mu_{\varphi_2}, \tau_{\varphi_2}, \mu_{p_2}, \tau_{p_2}, \mu_h, \tau_h, \mu_{G_b}, \tau_{G_b}, \mu_{I_b}, \tau_{I_b}, \mu_{G_0}, \tau_{G_0}, \mu_{I_0}, \tau_{I_0}).$$

We do not assume that the individual precisions $\tau_{gj}, \tau_{xj}, \tau_{ij}, \tau_{g_j^o}$ and $\tau_{i_j^o}$ each arises from a common population distribution, in order to reduce the number of unknown parameters and allow for a flexibility in the process errors. Furthermore these precisions are not of particular interest.

The population based minimal model can then be illustrated by the directed graph in Figure 5, where the posterior of the unobserved quantities given the observed data is given as

$$p(\Upsilon, \mathbf{\Omega}, \mathbf{\Psi} \,|\, \mathbf{\Phi}) \propto p(\Upsilon) \prod_{j=1}^{L} p(\Phi_j \,|\, \Omega_j, \Psi_j) p(\Psi_j \,|\, \Omega_j) p(\Omega_j \,|\, \Upsilon), \tag{10}$$

with $\mathbf{\Omega} = (\Omega_1, \ldots, \Omega_L)$, $\mathbf{\Psi} = (\Psi_1, \ldots, \Psi_L)$ and $\mathbf{\Phi} = (\Phi_1, \ldots, \Phi_L)$. The parent-child distributions, $p(\Psi_j \,|\, \Omega_j)$ and $p(\Phi_j \,|\, \Omega_j, \Psi_j)$, are already specified in (7) and (8), whereas the parent-child distribution $p(\Omega_j \,|\, \Upsilon)$ can be found using the recursive factorization of the directed graphical model as

$$\begin{aligned} p(\Omega_j \,|\, \Upsilon) \;=\; & p(S_{Gj} \,|\, \mu_{S_G}, \tau_{S_G}) p(S_{Ij} \,|\, \mu_{S_I}, \tau_{S_I}) p(\varphi_{1j} \,|\, \mu_{\varphi_1}, \tau_{\varphi_1}) p(\varphi_{2j} \,|\, \mu_{\varphi_2}, \tau_{\varphi_2}) p(p_{2j} \,|\, \mu_{p_2}, \tau_{p_2}) \\ & \times p(h_j \,|\, \mu_h, \tau_h) p(G_{bj} \,|\, \mu_{G_b}, \tau_{G_b}) p(I_{bj} \,|\, \mu_{I_b}, \tau_{I_b}) p(G_{0j} \,|\, \mu_{G_0}, \tau_{G_0}) p(I_{0j} \,|\, \mu_{I_0}, \tau_{I_0}) \\ & \times p(\tau_{gj}) p(\tau_{xj}) p(\tau_{ij}) p(\tau_{g^o j}) p(\tau_{i^o j}) \end{aligned}$$

corresponding to the density specified in (9) conditioned on the population parameters in $\Upsilon$. Concerning the prior for $\Upsilon$, $p(\Upsilon)$, we assume that the elements of $\Upsilon$ are independent

and that each of the mean parameters $\mu_{S_G}, \mu_{S_I}, \ldots, \mu_{I_0}$ are log-normally distributed, and the precisions $\tau_{S_G}, \tau_{S_I}, \ldots, \tau_{I_0}$ each has a uniform shrinkage prior, all with large variances, as we do not want the posterior to be dominated by the prior. Hereby the prior density is given as

$$\begin{aligned} p(\Upsilon) &= p(\mu_{S_G})p(\tau_{S_G})p(\mu_{S_I})p(\tau_{S_I})p(\mu_{\varphi_1})p(\tau_{\varphi_1})p(\mu_{\varphi_2})p(\tau_{\varphi_2})p(\mu_{p_2})p(\tau_{p_2}) \\ &\quad \times p(\mu_h)p(\tau_h)p(\mu_{G_b})p(\tau_{G_b})p(\mu_{I_b})p(\tau_{I_b})p(\mu_{G_0})p(\tau_{G_0})p(\mu_{I_0})p(\tau_{I_0}), \end{aligned}$$

where the densities for the mean parameters are log-normal whereas uniform shrinkage priors are used for all the precisions.

## 6. SIMULATION BASED INFERENCE

Statistical inference in the population based stochastic minimal model developed above may be done by constructing an MCMC simulation algorithm. This can be done in several ways.

One approach is to successively update each unobserved variable in $\mathbf{\Omega}$, $\mathbf{\Psi}$ and $\Upsilon$ conditioned on all the remaining variables. However, this approach eventually appears to be very inefficient due to bad mixing properties of the algorithm caused by the highly correlated quantities, i.e. even for very distant choices of parameter values, the two corresponding sets of latent processes may exhibit a surprising concordance. Nevertheless, these states must also be explored.

Another MCMC simulation scheme avoiding oscillating slowly between different states providing similar latent processes, would be to successively update the unobserved variables in blocks consisting of the parameters and the latent variables, respectively. This means updating the unobserved variables in three steps, where the first step is to draw a new state $\Upsilon'$ of the population parameters from a symmetric proposal distribution $q(\Upsilon; \Upsilon')$, where the acceptance probability given $\mathbf{\Omega}$ simply becomes

$$\alpha(\Upsilon; \Upsilon') = \min\left\{1, \frac{p(\Upsilon')\prod_{j=1}^{L} p(\Omega_j \mid \Upsilon')}{p(\Upsilon)\prod_{j=1}^{L} p(\Omega_j \mid \Upsilon)}\right\}.$$

Then afterwards the individual dependent parameters are updated by proposing in turn for subject $j$ a new state $\Omega_j'$ drawn from a symmetric proposal distribution $q(\Omega_j; \Omega_j')$, i.e. the acceptance probability given $(\Psi_j, \Phi_j, \Upsilon)$ becomes

$$\alpha(\Omega_j; \Omega_j') = \min\left\{1, \frac{p(\Phi_j \mid \Omega_j', \Psi_j)p(\Psi_j \mid \Omega_j')p(\Omega_j' \mid \Upsilon)}{p(\Phi_j \mid \Omega_j, \Psi_j)p(\Psi_j \mid \Omega_j)p(\Omega_j \mid \Upsilon)}\right\}.$$

Note how the updating of the individual subjects only depend upon the remaining subjects in the population through the introduced population parameters $\Upsilon$.

Finally the third step in the simulation algorithm would be to update $\Psi_j$ by proposing a new state $\Psi_j'$ from some symmetric proposal distribution $q(\Psi_j; \Psi_j')$, where the acceptance probability given $(\Omega_j, \Phi_j)$ becomes

$$\alpha(\Psi_j; \Psi_j') = \min\left\{1, \frac{p(\Phi_j \mid \Omega_j, \Psi_j')p(\Psi_j' \mid \Omega_j)}{p(\Phi_j \mid \Omega_j, \Psi_j)p(\Psi_j \mid \Omega_j)}\right\}.$$

This block-wise updating mechanism was expected to improve upon the oscillating behavior of the Markov chain, however, the resulting MCMC simulation algorithm was computationally very inefficient as the mechanism updating the latent processes $\Psi_j$ appeared to have acceptance probabilities very close to zero, i.e. the Markov chain did not freely traverse the state space as required.

Alternatively we replace step 2 and 3 by a simultaneous step for each subject $j$, i.e. we choose to update $\Omega_j$ by proposing a candidate $\Omega_j'$ from a symmetric proposal distribution $q(\Omega_j;\Omega_j')$ and then simulate $\Psi_j'$ from $p(\Psi_j \mid \Omega_j')$. This joint proposal is subsequently accepted with probability

$$\alpha(\Omega_j,\Psi_j;\Omega_j',\Psi_j') = \left\{1, \frac{p(\Phi_j \mid \Omega_j',\Psi_j')p(\Psi_j' \mid \Omega_j')p(\Omega_j' \mid \Upsilon)}{p(\Phi_j \mid \Omega_j,\Psi_j)p(\Psi_j \mid \Omega_j)p(\Omega_j \mid \Upsilon)}\right\}. \tag{11}$$

By updating $\Psi_j$ and $\Omega_j$ simultaneously we may suppress the strong inter-relationship between them, and hereby improve the simulation algorithm's mixing properties and its overall computational efficiency.

## 7. RESULTS

In this section we assess the performance of our approach to the minimal model by analyzing experimental IVGTT data from a population consisting of 19 healthy adults. The subjects were admitted to hospital and following an overnight fast, a glucose bolus (0.3 g glucose per kg bodyweight) were administered at time zero. Prior to the bolus, blood samples were taken at time -30, -15 and 0 for determination of the basal levels of insulin and glucose in the blood plasma. However, in the approach presented here, these measurements will not be exploited further. Blood samples were taken at 2, 4, 6, 8, 10, 15, 20, 25, 30, 40, 50, 60, 75, 90, 105, 120, 150, 180, 210 and 240 minutes for measurement of plasma glucose and insulin concentrations. For 12 of the 19 subjects, data were only gathered for 90 minutes. Figure 6 displays the recorded plasma glucose and plasma insulin concentrations on a log-scale during the IVGTT for all 19 healthy subjects.

The population based Bayesian approach to the minimal model was implemented according to the prescriptions given in Section 6. However, a few practical implementational issues must be addressed here. The logarithmic transformation brings the latent processes on the same scale and consequently we will model the random deviations from the model by a single unified Brownian motion $B$, i.e. we denote by $\tau$ the common precision $\tau = \tau_g = \tau_x = \tau_i$ in all three latent processes in $\Psi$. Furthermore, in order to ensure that the posterior distribution in (10) is dominated by the likelihood, we adopt a prior distribution with only little information, i.e. we center our a priori knowledge upon reported normal ranges for the parameters of interest though using large variances. Finally, for the glucose bolus administered at time zero to be fully distributed within the glycemic system, we only consider observations from time $t = 2$ minutes and onwards implying that the blood sample taken at time $t = 0$ is discarded.

Having established the prior and the remaining parent-child distributions in the posterior, we need only to specify the proposal distribution in order to establish a Markov chain with the posterior as target distribution. As discussed in Section 6 we update the entire state vector in order to ensure good mixing properties in the chain. Thus given the current state $(\Omega_j,\Psi_j)$

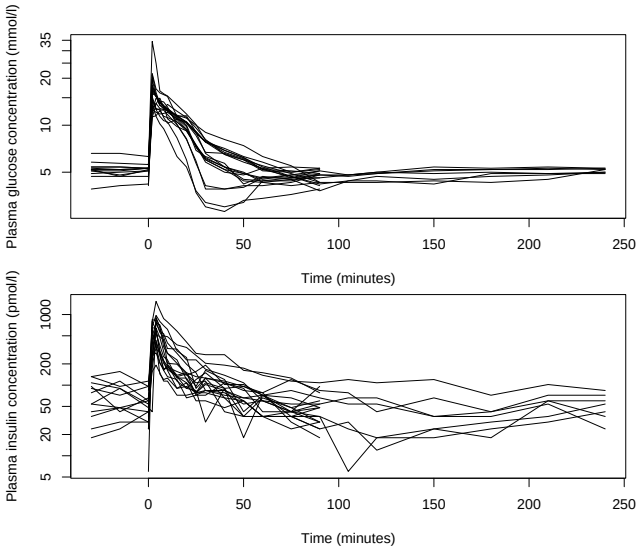



Figure 6. Plasma glucose concentrations and plasma insulin concentrations for 19 healthy subjects during an IVGTT.

a multivariate proposal $(\Omega_j', \Psi_j')$ has to be generated. For the $j$th subject, we first propose $\Omega_j'$ by drawing a random vector $\nu$ following a normal distribution, $\nu \sim \mathcal{N}(0, \Sigma)$, and then let $\Omega' = \Omega + \nu$. Note, in order to maintain a reversible Markov chain, this proposal distribution has to be independent of the particular subject under update. We therefore suggest making the elements of $\nu$ independent of each other, i.e. $\nu \sim \mathcal{N}(0, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_{10}^2))$. Secondly we simulate $\Psi_j$ from $p(\Psi_j \,|\, \Omega_j')$ according to (5). The acceptance probability for this update is subsequently given by (11).

In order to determine the proposal variances a large degree of pilot tuning is needed. We have automated this fine tuning by first running a chain for 100 000 iterations in which only one element of $\Omega$ is updated at a time, and then the corresponding proposal variance is scaled according to the acceptance probability. Next, a new Markov chain is run for another 100 000 iterations. Now all elements are updated simultaneously and again the variance is scaled according to the acceptance probability. Through this procedure, the Markov chain we have proposed here for assessing the normal metabolic portrait obtained satisfactory mixing properties and an overall acceptance probability of approximately 50 per cent.

The population parameter $\Upsilon$ is a stochastic entity which also require updating. We construct a symmetric proposal distribution $q(\Upsilon; \Upsilon')$ in a semi-automated way. Initially we use the proposal distribution for the individual metabolic portrait as proposal distribution for the population metabolic portrait. An MCMC simulation is performed and through out the simulation the proposals are dynamically scaled to obtain an overall acceptance probability of 50 per cent.

In practice it is important to monitor the performance of the developed MCMC simulation algorithm so that we can be sure that the obtained chains have settled to the desired

distribution and, in addition, that we have produced a sufficiently long sample for statistical inference. There are at least three issues here. Firstly, the Markov chain we have proposed here for assessing the metabolic portrait was low level implemented in a stand alone C program and some experimental code-checking is called for. By letting $V(\Psi, \Omega) \equiv 0$ and $W(\Phi, \Omega, \Psi) \equiv 0$, so that the chain is entirely prior driven, we ran several simulations varying only the random seed. Since estimates from these different runs are similar, we conclude that all are sampling from the same stationary distribution. Secondly, the chain takes some time to settle to its stationary distribution and samples from this initial part of the chain is usually discarded. Various methods for testing the convergence of the chain have been proposed in the literature and we used the method of Geweke (1992) [22] implemented in the CODA package [23] for R/Splus to test for convergence of the Markov chain. Thirdly, once the chain has converged to its stationary distribution, the chain must be run long enough to obtain reliable inference. To ensure this, we propose using the Heidelberger and Welch's convergence criteria [24] which is also implemented in the CODA package. See Brooks and Gelman (1998) [25] and Brooks and Guidici (2000) [26] for a general review of diagnostic techniques for MCMC simulation. However, one must note that population based Bayesian analyses via MCMC techniques can be very cumbersome, as large populations increase simulation time dramatically. For convenience, we therefore instantiate a final run for 1 500 000 iterations. On a shared SunFire 280R with two 750MHz sun4u Ultra-SPARC-III processors with 4 GB memory this run takes approximately 100 hours.

The output from the MCMC simulation algorithm consists of samples from the population parameters and the individual samples from the subject dependent parameters. All outputs were closely inspected for convergence with the Geweke method and typically burn-in was reached at approximately 500 000 – 600 000 iterations. See Figure 7(a) and (b) for the trace plots concerning $\varphi_1$. We therefore assume that the Markov chain has settled to its stationary distribution at 750 000 iterations, i.e. the remaining 750 000 iterations are used for statistical inference.

Figure 7(e) gives the posterior density (thick line) of the population mean of $\varphi_1$ and seems rather narrow compared to the prior density used within the simulation algorithm (dashed line). The corresponding plots for the population means of $S_G$, $S_I$ and $\varphi_2$ are shown in Figure 7(c), (d) and (f). characteristics from the samples after burn-in.

It is straightforward with the population based Bayesian approach to the minimal model also to withdraw information on each individual within the population. Thus we can easily construct individual posterior densities, see Figure 8. From here it is evident that our approach to the minimal model provides non-negative estimates and credible intervals allowing for physiological meaningful interpretation.

Furthermore, we may extract credible intervals and mean estimates in order to make comparisons of the obtained individual metabolic portraits with the corresponding portraits obtained by use of the MINMOD computer programme [2]. Note, however, that MINMOD in its present form only allows for the estimation of $S_G$, $S_I$, $p_2$ and $G_0$. As only $S_G$ and $S_I$ are part of the metabolic portrait, we have compared these two sets of estimates. Figure 9 displays the 95 per cent posterior credible intervals obtained for $S_G$ and $S_I$ by the Bayesian approach (black boxes) contrasted with the corresponding MINMOD estimates (gray boxes). Note that MINMOD fails in more than 40 per cent of the population, i.e. predicts physiological meaningless confidence intervals or crashes during estimation (missing bars). We also note

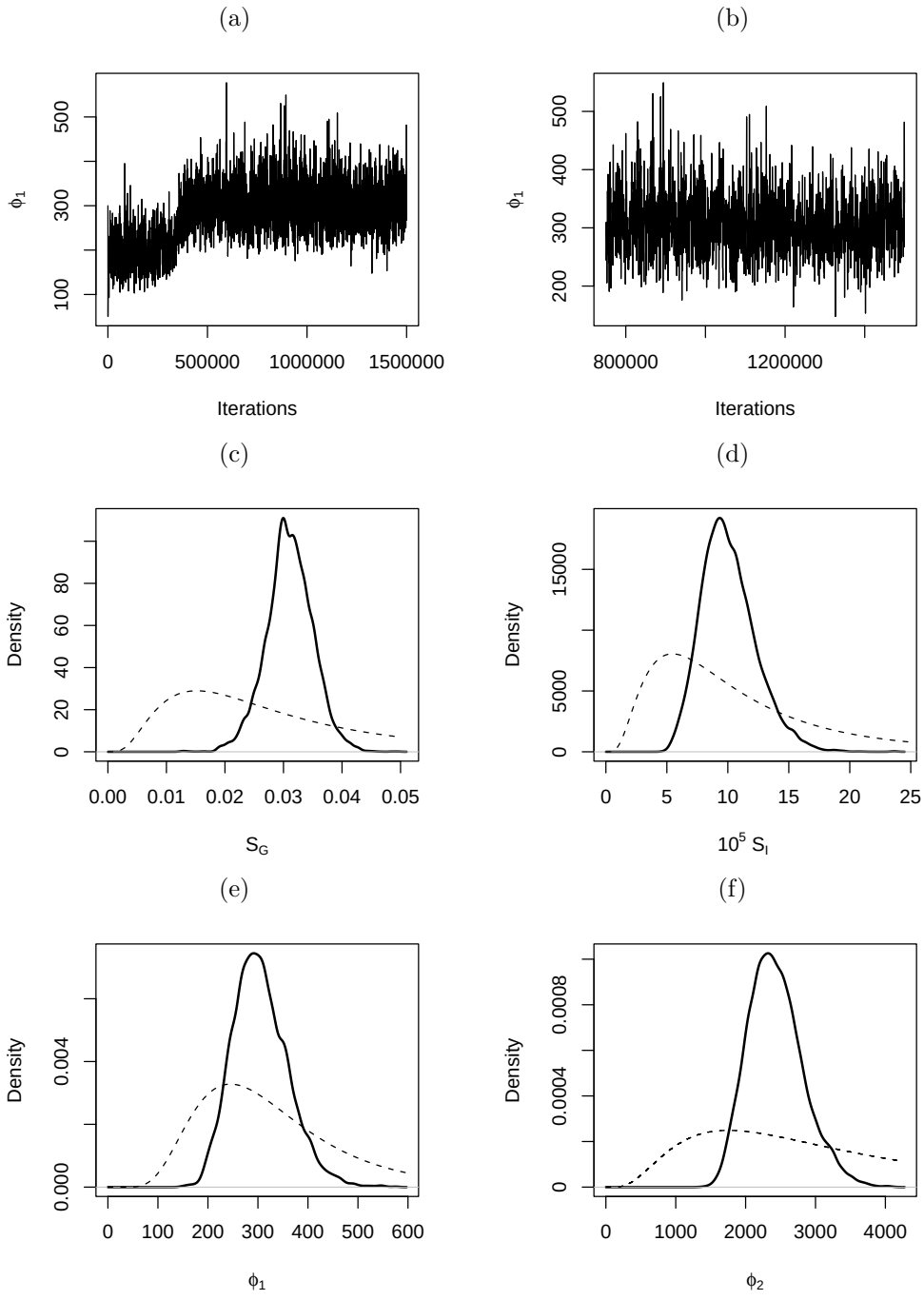



Figure 7. The trace plot obtained for $\varphi_1$ is shown in (a) together with the same trace plot after burn-in in (b). The corresponding posterior density (bold line) is given in (e) with the prior density superimposed (dashed line). The corresponding plots for $S_G$, $S_I$ and $\varphi_2$ are given in (c), (d) and (f).

Table I. The achieved posterior mean and standard deviation estimates together with 95 per cent credible intervals for the population means of the parameters in the minimal model.

| | | | 95% C.I. | |
|---|---|---|---|---|
| Parameter | Mean | St.d. | Lower | Upper |
| $\mu_{S_G}$ | 0.0310 | 0.0040 | 0.0230 | 0.0388 |
| $\mu_{S_I} \cdot 10^5$ | 10.0799 | 2.2227 | 6.3687 | 15.1568 |
| $\mu_{\varphi_1}$ | 305.0710 | 55.5230 | 209.2769 | 426.2393 |
| $\mu_{\varphi_2}$ | 2430.8615 | 391.8236 | 1766.5086 | 3288.0653 |
| $\mu_{G_b}$ | 5.8165 | 0.5902 | 4.8041 | 7.1257 |
| $\mu_{I_b}$ | 48.6063 | 7.8783 | 33.6677 | 65.0613 |
| $\mu_{G_0}$ | 26.7017 | 4.7257 | 19.5312 | 37.5714 |
| $\mu_{I_0}$ | 668.0657 | 72.8299 | 535.2292 | 822.6092 |
| $\mu_{p_2}$ | 0.1095 | 0.0538 | 0.0276 | 0.2336 |
| $\mu_h$ | 4.8838 | 0.1558 | 4.5907 | 5.2222 |

that the Bayesian credible intervals for $S_I$ generally appear wider than the corresponding confidence intervals obtained by MINMOD. This is particularly pronounced for the first 12 subjects, however, this might be explained by these subjects having data only for the first 90 minutes, but also by the fact that MINMOD is based upon the deterministic set of differential equations with the insulin as a forcing function. As $S_I$ in principle is governed by the active insulin, for which we have no observations, we will expect wider credible intervals than the confidence intervals provided by MINMOD. However, the increased uncertainty in $S_I$ implies that $S_G$ is more accurately determined.

An interesting feature of our approach to the minimal model is also revealed in Figure 9 as our approach generally predicts $S_G$ values smaller than the corresponding values predicted by MINMOD, whereas the exact opposite situation occurs under the estimation of $S_I$. As pointed out by Cobelli et al. (1999) [27], the traditional MINMOD approach has a tendency of underestimating $S_I$ and overestimating $S_G$ compared to the clamp study, which is considered as the gold standard. This indicates that our Bayesian approach might deal with these problems, but is a feature that needs further investigation.

## 8. DISCUSSION

In this work we adopt a directed graphical model as a powerful and flexible modeling framework for regularizing the ill-posed estimation problem possessed by coupling the three differential equations of the minimal model and considering them as stochastic entities. We have extended the minimal model to a population based hierarchical model including the minimal model for a single individual. The parameter estimation in this model of coupled stochastic differential equations is efficiently implemented in a Bayesian approach where posterior inference is performed through the use of Markov chain Monte Carlo techniques, particularly the Metropolis-Hastings algorithm. The parameter estimation hereby provides a quantitative assessment of the metabolic portrait both for a single individual, but also for a population, which is very useful in the prognosis and prevention of diabetes.

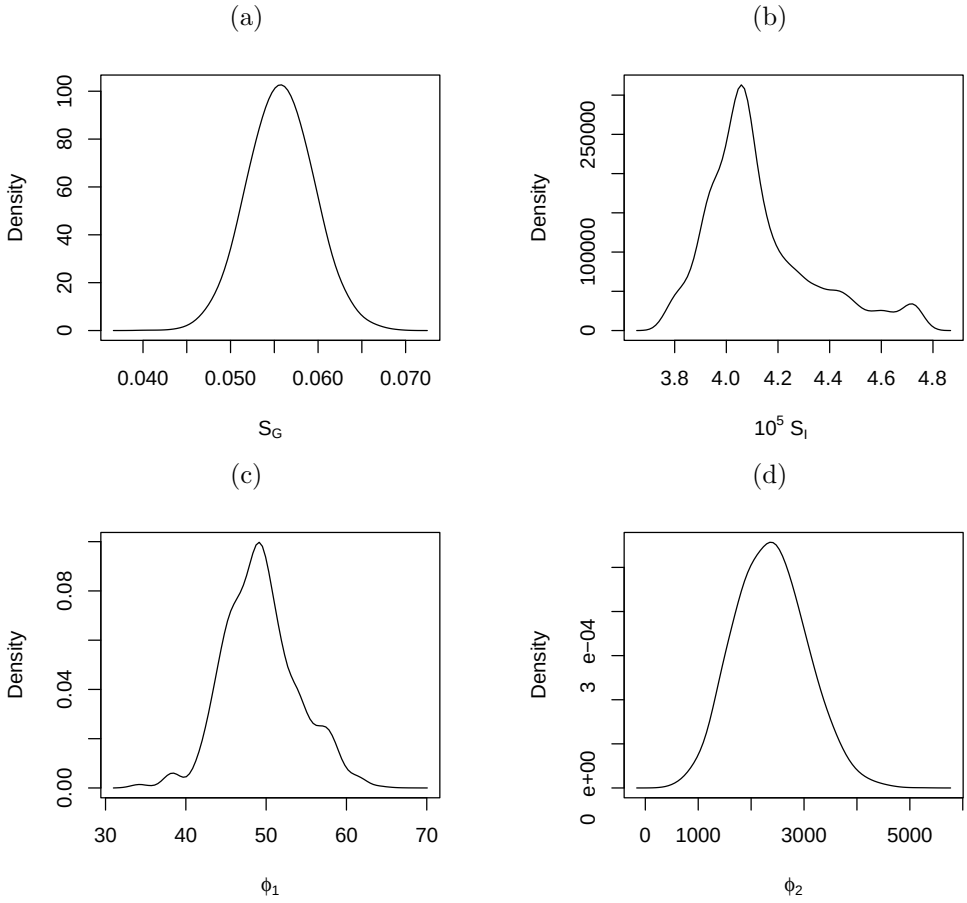



Figure 8. The posterior densities obtained during the Bayesian analysis for subject no. 19: (a) – (d) shows the posterior density of $S_G$, $S_I$, $\varphi_1$ and $\varphi_2$, respectively.

We consider a stochastic version of the minimal model, meaning that we include two types of errors in the model; noise on the process increments and noise on the observations. Such a stochastic approach to the minimal model, with distinction between observation error and process error, has to our knowledge not yet been studied in the literature. We believe it is more physiological correct also to include noise on the glucose and insulin processes, since the state of the tested subjects may also influence on the reaction in the processes. One might argue that one error term could sufficiently include both error terms. However, by adding noise with variance depending on the time interval $t_k - t_{k-1}$, we also account for errors arising from the discretization of the differential equations. Therefore we believe that having two error terms in our approach is the most suitable.

We consider all three differential equations of the minimal model, simultaneously, and hereby we obtain a coupled and computationally very complex system, which traditionally has been

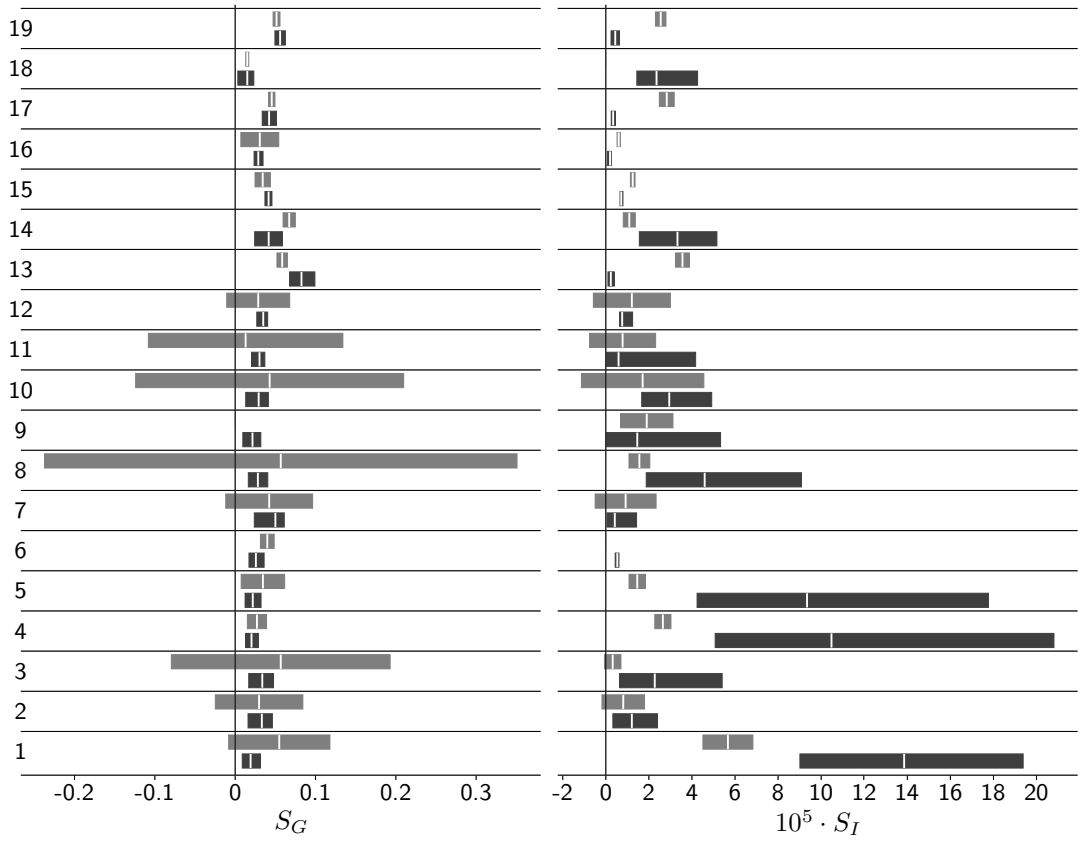



Figure 9. A comparison of the estimates obtained by the Bayesian approach and MINMOD. Black boxes denote 95 per cent credible intervals for the Bayesian approach, whereas gray boxes denote 95 per cent confidence intervals obtained for the same subject by MINMOD. The white line dividing the boxes indicates the mean. Numbers to the left indicate subject numbers.

analyzed in a more simple way, where the insulin part has been treated as known. This approach has more or less been the standard approach for analyzing the minimal model since it was proposed in the late seventies. Bergman et al. (1979) [28] claim that ignoring the noise on the insulin samples, and linearly interpolating over them, does not influence the final estimates of the parameters. However, as pointed out in Cobelli et al. (1999) [27], this traditional approach has a tendency of underestimating $S_I$ and overestimating $S_G$. Therefore it would have been interesting to compare our results with the corresponding clamp studies on the same subjects, but this has not been possible with the current data set. However, we have been able to compare our results with the traditional approach by analyzing the individual data in the MINMOD program. Compared to MINMOD there is a tendency of higher estimates of $S_I$ and smaller estimate of $S_G$ in our Bayesian approach, which indicates that our method might cope with the under- and overestimating problems occurring in the traditional approach. This is an issue that needs further investigation, for example on a larger data set for Type II diabetic

subjects where the under- and overestimation is even more pronounced.

We analyze the model in a Bayesian approach to regularize the ill-posed estimation problem of the coupled model by using a prior distribution, where the implausible parameters automatically are penalized. Hereby we cope with the problems in the traditional non-linear least squares method where unrealistic results are obtained ($S_I$ equals zero and negative confidence intervals for strictly positive parameters). Bayesian approaches to these problems have earlier been investigated in Pillonetto et al. (2002) [4], Vicini and Cobelli (2001) [6] and Agbaje et al. (2003) [7], but in all these approaches the insulin is treated as known and the model is not allowed to deviate from the differential equations as we do. Therefore our approach can be seen as an extension and more general approach compared to the other methods. Furthermore we allow all quantities of the model to be random, e.g. $G_b$ and $I_b$, and simulate values of these. In the other Bayesian approaches they are fixed at constant values and not simulated, even though the results are very depend on these values. In Bayesian statistics it is fundamental to consider all quantities as random, thus, it seems more natural to simulate values of them rather than fixing them. There are other advantages with the Bayesian approach, namely the computational tools (such as MCMC methods) are available, which allow for construction of suitably complex models without the need for simplifying assumptions. Therefore we are actually able to consider all quantities of the model and all three differential equations simultaneously as described above. Computationally the Bayesian approach is more demanding compared to the traditional method. However, the metabolic portrait is often estimated in large-scale and time-consuming epidemiological studies, where computations of 100 hours seem to vanish. Nevertheless, by constructing proposal distributions based on the gradient, much more efficient Monte Carlo methods may be exploited.

We establish a hierarchical model collecting the single individuals jointly together in a population model. Opposite to the traditional approach the individuals are able to share information about each other in this model through the common population parameters, whereby less informative a priori knowledge is required in the Bayesian analysis compared to earlier studies, [29], where only one single individual is included. Furthermore we experience that fewer simulations are needed in order to achieve convergence, also compared to single individual models. Vicini and Cobelli (2001) [6] and Agbaje et al. (2003) [7] also take a population based Bayesian approach to the minimal model, but again, as pointed out above, our method is an extension of their approach.

The main result of the paper is development of a new and general framework for parameter estimation in stochastic differential equations models with highly correlated parameters. First we re-parameterize (in this particular case we log-transform) the model, such that the differential equations are on the same scale, then we approximate the differential equations by discretization and adding noise (i.e. consider the processes as Itô processes), then we specify the obtained model as a directed graphical model, and perform inference in this model in a Bayesian approach where posterior sampling is based on MCMC methods and blocking updates. Hereby we are able to estimate highly correlated parameters in complex stochastic systems. It looks like a promising framework and seems like a general method that can be used in similar stochastic differential equation models, however further investigation and experiences with the method are needed before a general approach can be proposed.

ACKNOWLEDGEMENT

This work has partly been funded by Novo Nordisk A/S.

REFERENCES

[1] Bergman RN, Ider YZ, Bowden CR, Cobelli C. Quantitative estimation of insulin sensitivity. *American Journal of Physiology*, 236(6):E667 – E677, 1979.

[2] Pacini G, Bergman RN. MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test. *Computer Methods and Programs in Biomedicine*, 23:113 – 122, 1986.

[3] Martin BC, Warram JH, Krolewski AS, Bergman RN, Soeldner JS, Kahn CR. Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study. *The Lancet*, 340:925 –929, 1992.

[4] Pillonetto G, Sparacino G, Magni P, Bellazzi R, Cobelli C. Minimal model $S_I = 0$ problem in NIDDM subjects: nonzero bayesian estimates with credible intervals. *American Journal of Physiology - Endocrinology and Metabolism*, 282:E565 – E573, 2002.

[5] de Gaetano A, Arino O. Mathematical modelling of the intravenous glucose tolerance test. *Journal of Mathematical Biology*, 40:136 – 168, 2000.

[6] Vicini P, Cobelli C. The iterative two-stage population approach to IVGTT minimal modeling: improved precision with reduced sampling. *American Journal of Physiology - Endocrinology and Metabolism*, 280:E179 – E186, 2001.

[7] Agbaje OF, Luzio SD, Albarrak AIS, Lunn DJ, Owens DR, Hovorka R. Bayesian hierarchical approach to estimate insulin sensitivity by minimal model. *Clinical Science*, 105:551 – 560, 2003.

[8] Lauritzen SL. *Graphical Models*. Clarendon Press, Oxford, UK, 1996.

[9] Robert CP, Casella G. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 1999.

[10] Bergman RN, Phillips LS, Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man. *Journal of Clinical Investigation*, 68:1456 – 1467, 1981.

[11] Toffolo G, Bergman RN, Finegood DT, Bowden CR, Cobelli C. Quantitative estimation of $\beta$ cell sensitivity to glucose in the intact organism: a minimal model of insulin kinetics in the dog. *Diabetes*, 29:979 – 990, 1980.

[12] Saad MF, Anderson RL, Laws A, Watanabe RM, Kades WW, Chen YDI, Sands RE, Pei D, Savage PJ, Bergman RN. A comparison between the minimal model and the glucose clamp in assessment of insulin sensitivity across the spectrum of glucose tolerance. *Diabetes*, 43:1114 – 1121, 1994.

[13] Barrett PHR, Bell BM, Cobelli C, Golde H, Schumitzky A, Vicini P, Foster D. SAAM II: simulation, analysis and modeling software for tracer and pharmacokinetic studies. *Metabolism*, 47:484 – 492, 1998.

[14] D'Argenio D, Schumitzky A. ADAPT II Users Guide: Pharmacokinetic/Pharmacodynamic Systems Analysis Software. Technical report, Los Angeles, CA: Biomedical Simulations Resource, University of Southern California, 1997.

[15] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087 – 1092, 1953.

[16] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97 – 109, 1970.

[17] Brooks SP. Markov chain Monte Carlo method and its application. *The Statistician*, 47:69 – 100, 1998.

[18] Casella G, George EI. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

[19] Spiegelhalter DJ. Bayesian graphical modelling: a case-study in monitoring health outcomes. *Applied Statistics*, 47:115–133, 1998.

[20] Lunn DJ, Thomas A, Best N, Spiegelhalter DJ. WinBUGS - a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337, 2000.

[21] Kloeden PE, Platen E. *Numerical Solution of Stochastic Differential Equations*. Springer Verlag, 1999.

[22] Geweke J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J.M. Bernado, J.O. Berger, A.P. Dawid, A.F.M. Smith, editors, *Bayesian Statistics 4*, Oxford, UK, 1992. Clarendon Press.

[23] Best N, Cowles MK, Vines SK. *CODA: convergence diagnostics and output analysis software for Gibbs sampling output*. Version 0.6-1, MRC Biostatistics Unit, Cambridge, UK, 2003.

[24] Heidelberger P, Welch PD. A spectral method for confidence interval generation and run length control in simulations. *Communications of the Association for Computing Machinery*, 24:233 – 245, 1981.

[25] Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational Graphical Statistics*, 7:434 – 455, 1998.

[26] Brooks SP, Guidici P. MCMC convergence assessment via two-way ANOVA. *Journal of Computational Graphical Statistics*, 9:266 – 285, 2000.

[27] Cobelli C, Caumo A, Omenetto M. Minimal model $S_G$ overestimation and $S_I$ underestimation: improved accuracy by a Bayesian two-compartment model. *American Journal of Physiology - Endocrinology and Metabolism*, 277:E481 – E488, 1999.

[28] Bergman RN, Bortolan G, Cobelli C, Toffolo G. Iddentification of a minimal model of glucose disappearance for estimating insulin sensitivity. In *Proceedings of IFAC symposium identifcation and system parameter estimation*, volume 2, pages 883–890, 1979.

[29] Andersen KE, Højbjerre M. A Bayesian approach to Bergman's minimal model. In C. M. Bishop, B. J. Frey, editors, *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, pages 236 – 243, 2003.