A Bayesian Approach to Bergman's Minimal Model

Kim E. Andersen
Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7G
DK-9220 Aalborg East
Denmark
emil@math.auc.dk

Malene Højbjerre
Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7G
DK-9220 Aalborg East
Denmark
malene@math.auc.dk

Abstract

The classical minimal model of glucose disposal was proposed as a powerful modeling approach to estimating the insulin sensitivity and the glucose effectiveness, which are very useful in the study of diabetes. The minimal model is a highly ill-posed inverse problem and most often the reconstruction of the glucose kinetics has been done by deterministic iterative numerical algorithms. However, these algorithms do not consider the severe ill-posedness inherent in the minimal model and may only be efficient when a good initial estimate is provided. In this work we adopt graphical models as a powerful and flexible modeling framework for regularizing the problem and thereby allow for estimation of the insulin sensitivity and glucose effectiveness. We illustrate how the reconstruction algorithm may be efficiently implemented in a Bayesian approach where posterior sampling is made through the use of Markov chain Monte Carlo techniques. We demonstrate the method on simulated data.

1 INTRODUCTION

Diabetes is associated with a large number of abnormalities in insulin metabolism, ranging from an absolute deficiency to a combination of deficiency and resistance, causing an inability to dispose glucose from the blood stream. Three factors, referred to as *The Metabolic Portrait* (Pacini and Bergman, 1986), play an important role for glucose disposal

Insulin sensitivity: the capability of insulin to increase glucose disposal to muscles, liver and adipose tissue.

Glucose effectiveness: the ability of glucose to enhance its own disposal at basal insulin level.

Pancreatic responsiveness: the ability of the pancreatic β-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 is possible by the *The Minimal Model* (Bergman et al., 1979), and may improve classification, prognosis and therapy of the disease (Martin et al., 1992).

The minimal model is based on an *Intravenous Glucose Tolerance Test* (IVGTT), where glucose and insulin concentrations in plasma are sampled after an intravenous glucose injection. In the minimal model the glucose and insulin kinetics are described by two components, where the parameters traditionally have been estimated separately within each component. The glucose-insulin system is an integrated system and coupling of the components to obtain a unified model seems appropriate. However, this leads to a highly ill-posed inverse 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.

In the Bayesian approach for solving ill-posed inverse problems presented here, the available priori information is used to construct an efficient representation of the unknown quantities to be recovered. Thus we combine the components to obtain a unified model, and by adopting a graphical model (Lauritzen, 1996), we estimate the parameters in a Bayesian approach, where posterior sampling is performed by Markov chain Monte Carlo (MCMC) methods.

2 BERGMAN'S MINIMAL MODEL

In an IVGTT study a dose of glucose (usually 0.3 gr of glucose per kg body weight) is administered intravenously over a 60 seconds period to overnight-fasted subjects, and subsequently the glucose and insulin concentrations in plasma are frequently sampled (usually 30 times) over a period of 180 minutes. Data from a normal glucose tolerant individual is shown in Figure 1 (Pacini and Bergman, 1986).

The intravenous glucose dose immediately elevates the glu-

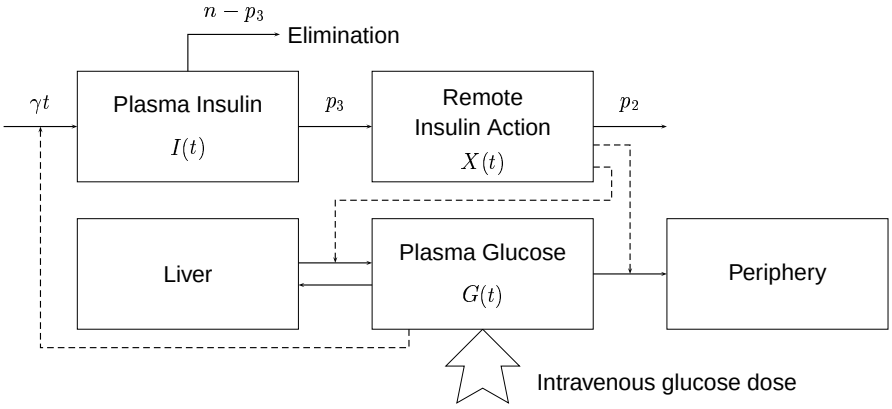

Figure 2: 'Bergman's Minimal Model' describing the glucose and insulin kinetics in an IVGTT study.

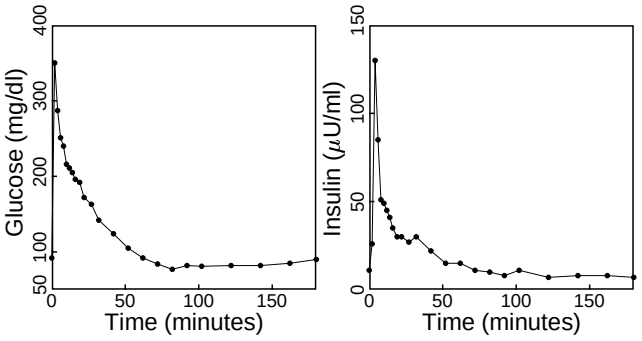

Figure 1: Glucose and insulin concentrations in plasma frequently sampled over 180 minutes after an intravenous glucose injection given to a normal glucose tolerant individual.

cose concentration in plasma forcing the pancreatic β-cells to secrete insulin. The insulin in plasma is hereby increased, and the glucose uptake in muscles, liver and tissue is raised by the remote insulin in action. This lowers the glucose concentration in plasma, implying the β-cells to secrete less insulin, from which a feedback effect arises. This integrated glucose-insulin system is illustrated by the compartment model in Figure 2, which can be described by the following non-linearly coupled system of differential equations (see e.g. Gaetano and Arino (2000) for details)

$$\begin{aligned}
\dot{G}(t) &= -p_1(G(t) - G_b) - X(t)G(t), & G(0) &= G_0,\\
\dot{X}(t) &= -p_2X(t) + p_3(I(t) - I_b), & X(0) &= 0,\\
\dot{I}(t) &= -n(I(t) - I_b) + \gamma(G(t) - h)^+t, & I(0) &= I_0,
\end{aligned}$$

where $t = 0$ is the glucose injection time, $^+$ denotes positive reflection and

$G(t)$: the glucose concentration in plasma [mg/dl] at time t [min].

$I(t)$: the insulin concentration in plasma [μU/ml] at time t [min].

$X(t)$: the insulin's effect on the net glucose disappearance (remote insulin in action) [min^{-1}].

G_b: the basal preinjection level of glucose [mg/dl].

I_b: the basal preinjection level of insulin [μU/ml].

p_1: the insulin-independent rate constant of glucose uptake in muscles, liver and adipose tissue [min^{-1}].

p_2: the rate for decrease in tissue glucose uptake ability [min^{-1}].

p_3: the insulin-dependent increase in glucose uptake ability in tissue per unit of insulin concentration above I_b [$\text{min}^{-2}(\mu\text{U/ml})^{-1}$].

n: the first order decay rate for insulin in plasma [min^{-1}].

h: the threshold value of glucose [mg/dl] above which the pancreatic β-cells release insulin.

γ: the rate of the pancreatic β-cells' release of insulin after the glucose injection and with glucose concentration above h [$(\mu\text{U/ml}\,\text{min}^{-2}(\text{mg/dl})^{-1}$].

G_0: the theoretical glucose concentration in plasma [mg/dl] at time 0.

I_0: the theoretical insulin concentration in plasma [μU/ml] at time 0.

The metabolic portrait of a single individual is then determined by the following parameters

Insulin sensitivity:	$S_I = \dfrac{p_3}{p_2}$,
Glucose effectiveness:	$S_G = p_1$,
Pancreatic responsiveness:	$\phi_1 = \dfrac{I_{\max} - I_b}{n(G_0 - G_b)}$,
	$\phi_2 = \gamma \times 10^4$,

where $I_{\max}$ is the maximum value of insulin in plasma. Note that S_I is measured in $(\mu\text{U/ml})^{-1}$ per minute, S_G in min^{-1} and ϕ_1 in $\text{min}^{-1}\mu\text{U/ml}$ per mg/dl.

The model parameters have usually been estimated by a non-linear weighted least squares estimation technique in a two-step procedure, where the parameters in $\dot{G}$ and $\dot{X}$ are estimated using insulin as a forcing function and then the parameter in $\dot{I}$ are estimated using glucose as a forcing function. However, the glucose-insulin system is an integrated system, and must be considered as a whole.

3 BERGMAN'S MINIMAL MODEL AS A STATISTICAL MODEL

The glucose and insulin concentrations are positive, and experience shows that the variability in the samples increases with the mean. Therefore we assume that both $G(t)$ and $I(t)$ are log-normally distributed and introduce

$$\begin{aligned} g(t) &= \log G(t) \quad \Rightarrow \quad \dot{g}(t) = \frac{\dot{G}(t)}{G(t)}, \\ x(t) &= \log X(t) \quad \Rightarrow \quad \dot{x}(t) = \frac{\dot{X}(t)}{X(t)}, \\ i(t) &= \log I(t) \quad \Rightarrow \quad \dot{i}(t) = \frac{\dot{I}(t)}{I(t)}, \end{aligned}$$

where $g(t)$ and $i(t)$ are normally distributed. This logarithmic transformation implies that the minimal model can be rewritten as

$$\begin{aligned} \dot{g}(t) &= -p_1(1 - G_b\, e^{-g(t)}) - e^{x(t)}, \\ \dot{x}(t) &= -p_2(1 - S_I(e^{i(t)} - I_b)e^{-x(t)}), \\ \dot{i}(t) &= -n(1 - e^{-i(t)} I_b) + e^{-i(t)}\gamma(e^{g(t)} - h)^+ t, \end{aligned} \tag{1}$$

subject to the initial conditions

$$\begin{aligned} g(0) &= \log(G_0), \\ x(0) &\to -\infty, \\ i(0) &= \log(I_0). \end{aligned}$$

Hereby the minimal model has no unit of measurement, and the three processes can be re-parameterized, such that they are on the same scale.

Glucose disposal described by the deterministic minimal model in (1) may, however, not comply with the actual 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

$$\begin{aligned} dg(t) &= \big(-p_1(1 - G_b e^{-g(t)}) - e^{x(t)}\big)dt + \tau_g^{-1/2} dB^g(t), \\ dx(t) &= \big(-p_2(1 - S_I(e^{i(t)} - I_b)e^{-x(t)})\big)dt + \tau_x^{-1/2} dB^x(t), \\ di(t) &= \big(-n(1 - e^{-i(t)} I_b) + e^{-i(t)}\gamma(e^{g(t)} - h)^+ t\big)dt \\ &\qquad + \tau_i^{-1/2} dB^i(t), \end{aligned}$$

where τ_g, τ_x and τ_i denote the reciprocal variances (the so-called precisions) of the Brownian motions accounting for model deviations. 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

$$\begin{aligned} g(t + \Delta t) - g(t) = \int_t^{t+\Delta t} & (-p_1(1 - G_b e^{-g(t)}) - e^{x(t)})dt \\ & + \tau_g^{-1/2}(B^g(t + \Delta t) - B^g(t)). \end{aligned}$$

The involved unknown integral is approximated by the product between its width and its left end point, that is

$$g(t+\Delta t) = g(t) - \Delta t\big(p_1(1 - G_b e^{-g(t)}) - e^{x(t)}\big) + \epsilon_t^g(\Delta t),$$

where the random process $\epsilon_t^g(\Delta t) = \tau_g^{-1/2}(B^g(t + \Delta t) - B^g(t))$ is well-known to depend on Δt only and to follow a normal distribution with mean zero and variance $\tau_g^{-1}\Delta t$.

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

$$\begin{aligned} g_{t+\Delta t} &= f^g(g_t, x_t, \Delta t) + \epsilon_t^g(\Delta t), \\ x_{t+\Delta t} &= f^x(x_t, i_t, \Delta t) + \epsilon_t^x(\Delta t), \\ i_{t+\Delta t} &= f^i(i_t, g_t, \Delta t) + \epsilon_t^i(\Delta t), \end{aligned}$$

where

$$\begin{aligned} f^g(g_t, x_t, \Delta t) &= g_t - \Delta t\big(p_1(1 - G_b\, e^{-g_t}) + e^{x_t}\big), \\ f^x(x_t, i_t, \Delta t) &= x_t - \Delta t p_2\big(1 - S_I(e^{i_t} - I_b)e^{-x_t}\big), \\ f^i(i_t, g_t, \Delta t) &= i_t + \Delta t\big(-n(1 - e^{-i_t} I_b) \\ &\qquad + e^{-i_t}\gamma(e^{g_t} - h)^+ t\big), \end{aligned}$$

and we for notational convenience have suppressed the functional dependencies of the parameters $p_1, p_2, S_I, G_b, I_b, n, \gamma$ and h.

The conditional distributions for the processes $g_{t+\Delta t}$, $x_{t+\Delta t}$ and $i_{t+\Delta t}$ are given as

$$\begin{aligned} g_{t+\Delta t} \mid g_t, x_t, \tau_g &\sim \mathcal{N}(f^g(g_t, x_t, \Delta t), \tau_g^{-1}), \\ x_{t+\Delta t} \mid x_t, i_t, \tau_x &\sim \mathcal{N}(f^x(x_t, i_t, \Delta t), \tau_x^{-1}), \\ i_{t+\Delta t} \mid i_t, g_t, \tau_i &\sim \mathcal{N}(f^i(i_t, g_t, \Delta t), \tau_i^{-1}). \end{aligned} \tag{2}$$

The statistical dependencies in this model specifies a directed graphical model (Lauritzen, 1996) that can be illustrated by the directed acyclic graph in Figure 3, in which we have omitted the parameter vertices $p_1, p_2, S_I, G_b, I_b, n, \gamma, h, \tau_g, \tau_i$ and τ_x. In addition we have added the random variables

$$g_t^o = \log(G_t^o) \text{ and } i_t^o = \log(I_t^o),$$

where G_t^o and I_t^o are the random variables actually observed for specific values of t.

We model the measurement error on g_t^o and i_t^o by the random white noise processes; $\epsilon_t^{g^o}$ and $\epsilon_t^{i^o}$ with precisions τ_{g^o} and τ_{i^o}, i.e. the model assumptions for $g_{t+\Delta t}^o$ and $i_{t+\Delta t}^o$ are

$$\begin{aligned} g_{t+\Delta t}^o \mid g_{t+\Delta t}, \tau_{g^o} &\sim \mathcal{N}(g_{t+\Delta t}, \tau_{g^o}^{-1}), \\ i_{t+\Delta t}^o \mid i_{t+\Delta t}, \tau_{i^o} &\sim \mathcal{N}(i_{t+\Delta t}, \tau_{i^o}^{-1}). \end{aligned} \tag{3}$$

Notice that the mean structures of the logarithmically transformed observations are modelled by the underlying non-observable and hereby latent processes $g_{t+\Delta t}, x_{t+\Delta t}$ and $i_{t+\Delta t}$.

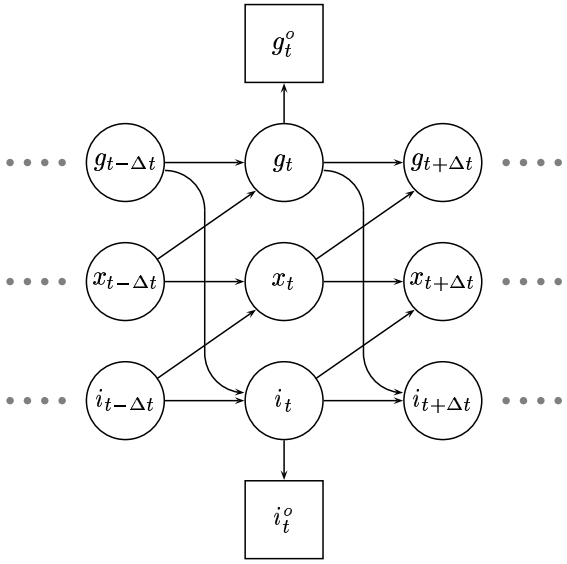

Figure 3: Directed acyclic graph illustrating the statistical dependencies for the g, x and i processes.

4 SIMULATION BASED INFERENCE

If we assume that $\Psi = \{g_t, x_t, i_t\}_{t\in\Lambda}$, where $\Lambda = \{\Delta t, 2\Delta t, \ldots, N\Delta t\}$, denotes the three latent processes in (2) and that the observed data are described by the model given in (3), then the statistical problem is to estimate the vector of unobserved parameters

$$\Theta = (p_1, p_2, S_I, \gamma, n, h, G_b, I_b, g_0, i_0, \tau_g, \tau_x, \tau_i, \tau_{g^o}, \tau_{i^o})$$

given the vector of the logarithmically transformed observations $\Phi = \{g_t^o, i_t^o\}_{t\in\mathcal{T}}$, where $\mathcal{T}$ denotes the set of observation times. For this we need to establish the posterior distribution $\pi(\Theta, \Psi \mid \Phi)$, which represents our beliefs about the feasible structures of (Θ, Ψ) after having observed the data Φ. Dividing all the quantities into subsets of data, Φ, latent processes, Ψ and parameters, Θ, the statistical dependencies in the model defined in (2) and (3) can be summarized by the simple directed acyclic graph in Figure 4, where it should be noticed that the time aspect depicted in Figure 3 has eventually vanished.

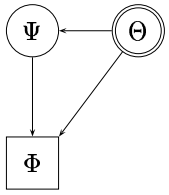

Figure 4: Directed acyclic graph illustrating the statistical relationship for the observed data, Φ, the latent variables, Ψ, and the parameters, Θ.

The recursive factorization of the directed graphical model implies that the posterior distribution factorizes as

$$\pi(\Theta, \Psi \mid \Phi) \propto p(\Theta)p(\Psi \mid \Theta)p(\Phi \mid \Theta, \Psi),$$

where $p(\Theta)$ represents our beliefs about the parameters before having observed any data and $p(\Psi \mid \Theta)$ and $p(\Phi \mid \Theta, \Psi)$ form the likelihood determined by (2) and (3). Performing inference about the parameters are reduced to the computational task of evaluating integrals over the state space of the latent variables and parameters, e.g.

$$\mathbb{E}_\pi(\Theta) = \iint_{\boldsymbol{\Theta}\times\boldsymbol{\Psi}} \Theta\, \pi(\Theta, \Psi \mid \Phi)\, d\Theta\, d\Psi.$$

Explicit evaluation of such integrals are impossible due to the huge state space, however, MCMC methods provide an approximative integration technique whereby marginal posterior means, for example, are estimated by using the sample mean from a representative series of random draws from the posterior distribution. These random draws are obtained by constructing an irreducible Markov chain $\{(\Theta_1, \Psi_1), (\Theta_2, \Psi_2), \ldots\}$ with state space $\boldsymbol{\Theta} \times \boldsymbol{\Psi}$ and with stationary distribution π. MCMC sampling was first introduced by Metropolis et al. (1953) and was subsequently adapted by Hastings (1970). Over the past ten years such methods have enjoyed widespread popularity within the statistical literature and there exist various standard techniques for constructing the necessary chains (see e.g. Brooks, 1998; Robert and Casella, 1999).

4.1 Metropolis–Hastings updates

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 configuration (Θ, Ψ), then we draw a new configuration (Θ', Ψ') from some proposal density $q(\Theta, \Psi; \Theta', \Psi')$. This proposal is then accepted with probability

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

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. $q(\Theta, \Psi; \Theta', \Psi') = q(\Theta', \Psi'; \Theta, \Psi)$, the acceptance function reduces to

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

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

4.2 Implementation

One approach for a MCMC simulation algorithm for the stochastic minimal model is successive updates of each unknown quantity given all the remaining quantities of the model. Due to the recursive factorization property of the directed graphical model the acceptance probabilities only depend locally on the updated quantity itself, its parents, its children and its childrens other parents, the so-called *Markov blanket*. This approach eventually appears to be very inefficient due to bad mixing properties of the algorithm caused by the highly correlated quantities.

Another approach would be to block the updates into Θ and Ψ, by first proposing a new state of the parameters Θ' drawn from a symmetric proposal distribution $q(\Theta;\Theta')$. Conditioned on Ψ and Φ the acceptance probability simply becomes

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

Then afterwards updating Ψ by proposing a new state Ψ' drawn from a symmetric proposal distribution $q(\Psi;\Psi')$, where the acceptance probability conditioned on Θ and Φ is

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

However, again the highly correlated quantities are expected to lead to a very inefficient MCMC simulation algorithm.

Alternatively we choose to update Θ by proposing a candidate Θ' from a symmetric proposal distribution $q(\Theta;\Theta')$ and then simulate Ψ' from $p(\Psi \mid \Theta')$. This proposal is subsequently accepted with probability

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

By updating Ψ and Θ simultaneously we may suppress the strong inter-relation between them, and thereby improve the simulation algorithm's mixing properties and overall efficiency.

In order to guarantee that the posterior distribution is dominated by the likelihood, we adopt a vague prior distribution $p(\Theta)$ on Θ. Thus we assume that the elements of Θ are independent and that each of the system parameters $p_1, p_2, S_I, n, \gamma, h, G_b, I_b, g_0$ and i_0 are log-normally distributed and that the precisions $\tau_g, \tau_x, \tau_i, \tau_{g^o}$ and τ_{i^o} are Gamma-distributed, all with large variances.

Consequently the prior density takes the simple form

$$\begin{aligned} p(\Theta) = {} & p(p_1)p(p_2)p(S_I)p(n)p(\gamma)p(h)p(G_b)p(I_b) \\ & \times p(g_0)p(i_0)p(\tau_g)p(\tau_x)p(\tau_i)p(\tau_{g^o})p(\tau_{i^o}), \end{aligned}$$

where the densities are either densities of a log-normal distribution or a gamma distribution.

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

$$\begin{aligned} p(\Psi \mid \Theta) = {} & \prod_{t\in\Lambda} p(g_t \mid g_{t-\Delta t}, x_{t-\Delta t}, \tau_g) \\ & \times p(x_t \mid x_{t-\Delta t}, i_{t-\Delta t}, \tau_x) \\ & \times p(i_t \mid i_{t-\Delta t}, g_{t-\Delta t}, \tau_i) \\ \propto {} & (\tau_g\tau_x\tau_i)^{(N+1)/2} \exp\{-V(\Psi,\Theta)\}, \end{aligned}$$

with

$$V(\Psi,\Theta) = \frac{1}{2}\sum_{t\in\Lambda} \tau_g(g_t - f_t^g)^2 + \tau_x(x_t - f_t^x)^2 + \tau_i(i_t - f_t^i)^2,$$

and

$$\begin{aligned} f_t^g &= f^g(g_{t-\Delta t}, x_{t-\Delta t}, \Delta t), \\ f_t^x &= f^x(x_{t-\Delta t}, i_{t-\Delta t}, \Delta t), \\ f_t^i &= f^i(i_{t-\Delta t}, g_{t-\Delta t}, \Delta t). \end{aligned}$$

Furthermore

$$\begin{aligned} p(\Phi \mid \Theta, \Psi) = {} & \prod_{t\in\mathcal{T}} p(g_t^o \mid g_{t-\Delta t}^o, \tau_{g^o}) p(i_t^o \mid i_{t-\Delta t}^o, \tau_{i^o}) \\ \propto {} & (\tau_{g^o}\tau_{i^o})^{|\mathcal{T}|/2} \exp\{-W(\Phi,\Theta,\Psi)\}, \end{aligned}$$

with

$$W(\Phi,\Theta,\Psi) = \frac{1}{2}\sum_{t\in\mathcal{T}} \tau_{g^o}(g_t^o - g_t)^2 + \tau_{i^o}(i_t^o - i_t)^2,$$

where $|\mathcal{T}|$ denotes the number of observations.

In practice the Metropolis–Hastings updating scheme described above can be used either to update the entire state vector or individual elements therein. Since proposals consisting of perturbations of the entire state vector tend to have correspondingly small acceptance probabilities, the proposed perturbations need to be kept at a suitable small level in order to achieve a satisfactory acceptance probability. However, this full component Metropolis–Hastings updating technique is known to lead to a Markov chain with slow convergence properties and therefore typical MCMC algorithms consist of a sequence of updates focusing upon each element of the state vector in turn. In Section 5, we shall combine this approach, known as single component Metropolis–Hastings, with updating of the entire vector to obtain a MCMC simulation algorithm with good convergence and mixing properties.

5 RESULTS

In this section we consider the performance of our approach on data simulated from the stochastic version of the minimal model derived in Section 3. The simulated data comes

Table 1: The choice of parameter values used for simulating the data depicted in Figure 5, prior assumptions, proposal distributions (numbers indicate the standard deviances of the proposals), initial values and the resulting posterior means, standard deviances and credible intervals for the unknown parameters.

			One component	Full component				95% C.I.	
Parameter	Truth	Prior	proposal (σ)	proposal (σ)	Initial	Mean	St.d.	Lower	Upper
p_1	0.0317	$\log\mathcal{N}(0.025, 1/36)$	1/1000	1/10000	0.0100	0.0316	0.0023	0.0271	0.0361
p_2	0.0123	$\log\mathcal{N}(0.01, 1/36)$	1/1000	1/10000	0.0300	0.0107	0.0018	0.0072	0.0142
S_I	0.0004	$\log\mathcal{N}(0.0005, 1/36)$	1/600000	1/800000	0.0001	0.0005	0.0001	0.0003	0.0007
γ	0.0039	$\log\mathcal{N}(0.0028, 1/36)$	1/400000	1/800000	0.0010	0.0042	0.0002	0.0038	0.0046
n	0.2659	$\log\mathcal{N}(0.2, 1/25)$	1/200	1/5000	0.1000	0.2640	0.0089	0.2465	0.2814
h	79.0353	$\log\mathcal{N}(100, 1/25)$	1/2	1/20	100.0000	80.2576	1.3670	77.5783	82.9370
G_0	291.2000	$\log\mathcal{N}(300, 1/64)$	1/20	1/40	200.0000	289.3832	3.4804	282.2594	296.6869
G_b	60.0000	$\log\mathcal{N}(60, 1/200)$	1/4	1/40	60.0000	62.2412	2.1782	57.9719	66.5105
I_0	364.8000	$\log\mathcal{N}(350, 1/64)$	1/20	1/80	400.0000	344.1208	16.1917	313.4896	377.7450
I_b	7.0000	$\log\mathcal{N}(7, 1/200)$	1/40	1/80	7.0000	6.7806	0.1125	6.5601	7.0011
τ	15000	$\Gamma(15, 0.0005)$	1000	200	20000	13266	2443	8491	18055
τ_{g^o}	500	$\Gamma(1, 0.001)$	25	–	50	463	115	236	690
τ_{i^o}	100	$\Gamma(1, 0.001)$	25	–	50	132	33	67	197

from a normal glucose tolerant individual with basal insulin level $I_b = 7\,\mu$U/ml and basal glucose level $G_b = 60$ mg/dl, i.e. we have chosen parameters according to the normal glucose tolerant population, see Table 1 for specific details.

The logarithmic transformation of the three latent processes brings them approximately on the same scale. We will therefore assume that the Brownian motions inherent within each process are independent and identically distributed. Subsequently we let $\tau = \tau_g = \tau_x = \tau_i$, whereby we have reduced the number of parameters to be estimated. The three latent processes $i(t), x(t)$ and $g(t)$ were simulated according to (2) with $\Delta t = 1$. The data available for statistical inference was then generated according to (3) with observations recorded at $t \in \mathcal{T}$, where $\mathcal{T} = \{0, 2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 16, 19, 22, 23, 24, 25, 27, 30, 35, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180\}$. The data obtained is depicted in Figure 5 together with the three latent processes. It is apparent that the injected glucose at $t = 0$ provokes an initial increment in glucose. By approximately 60 minutes glucose is normalized, and in the following two hours, a moderate undershoot is observed. Note how hyperglycemia induces an immediate peak in insulin followed by a second phase pancreatic responsiveness.

A MCMC simulation algorithm was constructed as described in Section 4. However, in order to efficiently explore the state space $\boldsymbol{\Theta}$ of Θ we propose combining the single component Metropolis–Hastings algorithm with the updating mechanism for the entire vector of parameters, i.e. the full component Metropolis–Hastings algorithm. Thus we may suggest more radical perturbations of the parameter of interest when using the single component Metropolis-Hastings algorithm whereas less radical perturbations are proposed when updating the entire vector. We base our a priori knowledge upon reported normal ranges for the parameters of interest, however, we may pre-record approximate basal lines in glucose and insulin prior to the exper-

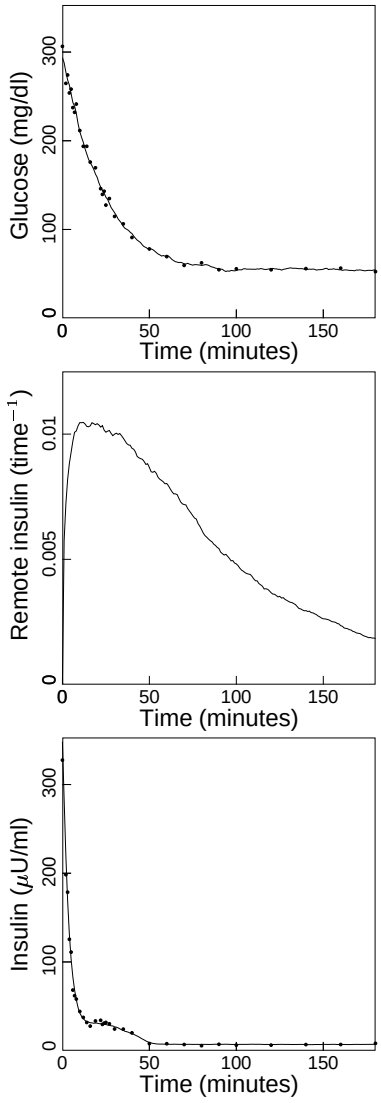

Figure 5: Simulated data. From top to bottom is shown: The glucose concentration; the remote insulin action; and the insulin concentration. Dots represent observed data and the lines represent the underlying processes.

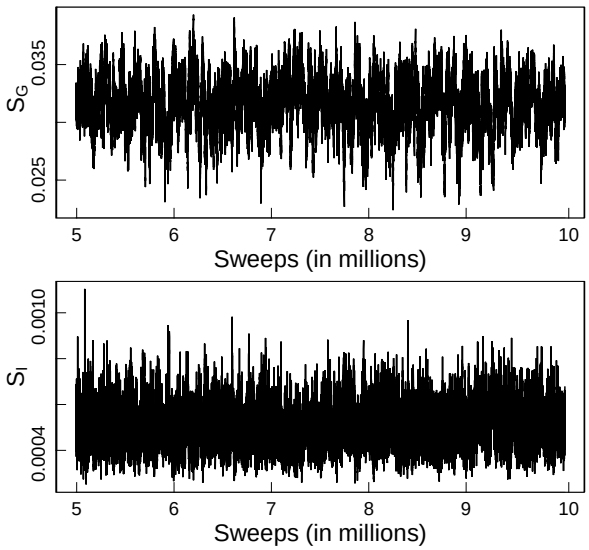

Figure 6: Trace plots from the Markov chain: From top to bottom: The trace plot of the glucose effectiveness S_G and the insulin sensitivity S_I.

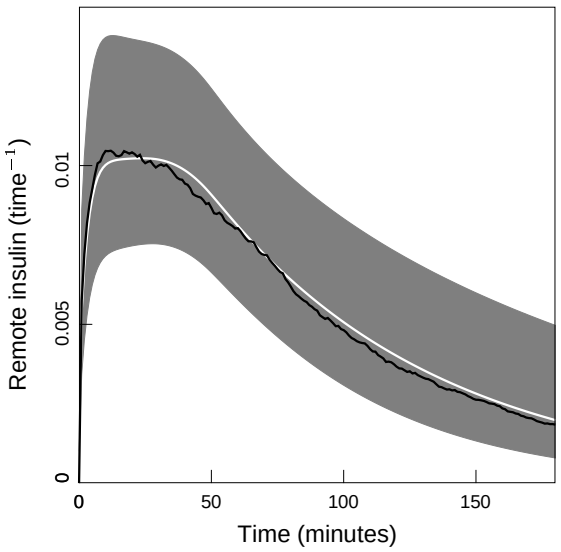

Figure 7: Posterior mean (white line) and 95% credible interval superimposed in gray for the remote insulin action (black line).

iment and will therefore model these two quantities with highly informative priors. Table 1 provides details on the prior information used within the MCMC simulation.

The MCMC simulation algorithm is based upon the random walk Metropolis–Hastings updating mechanism. Thus a candidate Θ' is generated by a symmetric perturbation of the current state Θ, i.e. for the full-component proposal distribution we will choose $\Theta' \sim \mathcal{N}(\Theta, \Sigma)$, where $\Sigma = \text{diag}(\sigma^2_{p_1}, \ldots, \sigma^2_\tau)$ is the diagonal covariance matrix specified by the standard deviances provided in Table 1. Likewise we update the single elements in Θ by random walk Metropolis–Hastings. Note that we update τ_{g° and τ_{i° separately.

The MCMC algorithm was initiated in an arbitrary vector Θ_0 far from the 'true' Θ, see Table 1 and ran for 10 000 000 iterations. The chain proposes single component updates 30 per cent of the time leading to an overall acceptance probability of 54.7 per cent. In Figure 6 we give the trace plots for the latter 5 000 000 samples obtained for the glucose effectiveness, S_G, and the insulin sensitivity, S_I.

The trace plots for the remaining parameters in Θ exhibit similar behavior and it is therefore apparent that the Markov chain exhibits rather excellent mixing properties. Inference about Θ is based upon the latter 5 000 000 samples. The posterior means and 95 per cent credible intervals for the parameters are given in Table 1, from which we may conclude that the Markov chain is sampling from the desired distribution. Note that all the 'true' values of the parameters are within the corresponding credible intervals, implying that the Bayesian approach to ill-posed inverse problems is a useful tool for regularization of the minimal model.

Furthermore, our approach also allows us to assess the uncertainty on the latent processes, e.g. the remote insulin action may be of special interest. The posterior mean of $X(t)$ is shown in Figure 7 together with its 95 per cent credible interval.

6 DISCUSSION

In this work we have adopted a graphical model as a powerful and flexible modeling framework for parameter estimation in general systems of coupled differential systems. In particular we have addressed the problem of regularizing the highly ill-posed inverse problem possessed by coupling the three differential equations in Bergman's minimal model for glucose and insulin kinetics. We have illustrated how the reconstruction may efficiently be implemented by a Bayesian graphical model where posterior sampling is performed through the use of Markov chain Monte Carlo techniques. Hereby we have made estimation of the highly correlated parameters in the minimal model possible even though we consider all three differential equations simultaneously, and we have provided a quantitative assessment of the metabolic portrait of a single individual, which is very useful in the prognosis and prevention of diabetes.

The method has been performed on simulated data and seems rather promising and very robust, but must also be proven useful on real field data. The Bayesian approach is based on specification of prior distributions on the parameters, which is a strength when prior information is available. We have used priors based on reported normal ranges, but with relatively large variances, requiring a prior sensitivity analysis to be considered.

The minimal model is developed for a single individual, and the extension to population modeling for the purpose of validating the normal metabolic portrait, is a potential of our method that is worth pursuing. However, this would probably require an even faster and more efficient simulations algorithm based upon e.g. the Metropolis adjusted Langevin algorithm.

Acknowledgements

This work has partly been funded by Novo Nordisk A/S. We are indebted to Bjarne Højgaard, Aalborg University, and Aage Vølund, Novo Nordisk A/S, for inspiring discussions and helpful comments.

References

Bergman, R. N., Ider, Y. Z., Bowden, C. R. and Cobelli, C. (1979). Quantitative estimation of insulin sensitivity, *American Journal of Physiology* **236**(6): E667 – E677.

Brooks, S. P. (1998). Markov chain Monte Carlo method and its application, *The Statistician* **47**: 69 – 100.

Gaetano, A. D. and Arino, O. (2000). Mathematical modelling of the intravenous glucose tolerance test, *Journal of Mathematical Biology* **40**: 136 – 168.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97 – 109.

Lauritzen, S. L. (1996). *Graphical Models*, Clarendon Press, Oxford, UK.

Martin, B. C., Warram, J. H., Krolewski, A. S., Bergman, R. N., Soeldner, J. S. and Kahn, C. R. (1992). 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.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087 – 1092.

Pacini, G. and Bergman, R. N. (1986). 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.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*, Springer-Verlag, New York.